SINGLE-CELL RNA SEQUENCING WITH WATERFALL REVEALS MOLECULAR CASCADES UNDERLYING ADULT HIPPOCAMPAL NEUROGENESIS

by
Jaehoon Shin, MD

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
February, 2016

Abstract

Somatic stem cells contribute to tissue ontogenesis, homeostasis, and regeneration through sequential processes. Systematic molecular analyses of stem cells and their development are challenging because classic approaches cannot resolve cellular heterogeneity or capture developmental dynamics. Here we provide a comprehensive resource of single-cell transcriptomes of adult hippocampal quiescent neural stem cells (qNSCs) and their immediate progeny. We further developed Waterfall, a bioinformatic suite, to statistically quantify singe-cell gene expression along de novo reconstructed continuous developmental trajectory. Our study reveals molecular signatures of qNSCs, characterized by high-niche signaling and low-protein translation capacities. Our analyses further delineate molecular cascades underlying adult qNSC activation and neurogenesis initiation, exemplified by decreased extrinsic signaling capacity, primed translational machinery, and switches in transcription factors, metabolism, and energy sources. Together, our study reveals the molecular continuum underlying adult neurogenesis and illustrates how Waterfall can be used for single-cell omics analyses of various continuous biological processes.

Advisor: Dr. Hongjun Song, PhD

Reader: Dr. Harry C. Dietz, MD

Acknowledgments

I would like to thank the faculty, trainees, and staff of Song laboratory at Johns Hopkins for their guidance and mentorship, specifically, Dr. Song, who inspired me to ask challenging questions and become an innovator in the field.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction[1]

[1] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

In discrete regions of the adult mammalian brain, quiescent neural stem cells (qNSCs) continuously generate new neurons through a recurrent process involving quiescent to active state transitions, cell cycle entry and neuronal fate specification (Ming and Song 2011). Understanding molecular mechanisms underlying adult NSC regulation and neurogenesis will not only advance our knowledge of neural development and plasticity, but enable new approaches for regenerative medicine and treatment of brain disorders. Mechanistic analysis of stem cell biology requires comprehensive quantification of molecular properties, such as gene expression. In contrast to traditional approaches targeting individual candidate genes, transcriptome profiling through RNA sequencing (RNA-seq) provides an unbiased and quantitative proxy for molecular features of cellular states. Such a blueprint may reveal unexpected features of NSC biology, generate hypotheses for functional analysis, and lead to novel strategies to manipulate neurogenesis processes.

Classic approaches for molecular characterization of somatic stem cell behavior use population-based readouts at a few time points along development, which faces two major challenges: resolving cellular heterogeneity and capturing developmental dynamics. Adult stem cells constitute a minor population within complex tissues, intermingled with their progeny at different developmental stages and supporting cells. They switch among different states, such as quiescence and activation (Li and Clevers 2010), and thus exhibit significant cellular and molecular differences even upon prospective isolation via fluorescence-activated cell sorting or genetic labeling (Lu, Neff et al. 2011, Codega, Silva-Vargas et al. 2014). Furthermore, snapshots of molecular composition at selected time points are not sufficient to understand the dynamic nature of stem cell development.

Single-cell RNA-seq generates gene expression profiles at the resolution of an individual cell and has thus far revealed molecular profiles of cell types that were not previously recognized at the population level (Stegle, Teichmann et al. 2015). Single-cell RNA-seq has not yet been widely adopted for adult somatic stem cell studies due to technical difficulties in obtaining individual stem cells from complex tissues. Further, the stochastic nature of gene expression in individual cells (Novick and Weiner 1957, Raj, Peskin et al. 2006, Muramoto, Cannon et al. 2012) may lead to overestimation of cellular heterogeneity and requires a new approach for statistical quantification. And for biological systems with only a few known markers, current approaches are

not sufficient to map out a developmental trajectory at high resolution from single-cell datasets (Bendall, Davis et al. 2014, Trapnell, Cacchiarelli et al. 2014).

Despite recent advances in acquiring snapshots of transcriptomes, epigenomes, and proteomes from individual cells, a remaining hurdle is the lack of methodology to identify molecular state transitions over a developmental continuum. Cells are destroyed during acquisition of omic data, so the same cell can't be tracked over time. Here we developed a conceptually different approach, analogous to the "shot-gun" method used in the human genome project characterized by parallel sequencing and bioinformatic reconstruction. We focused on the narrow time window of adult qNSC activation and neurogenesis initiation. Using the *Nestin-CFP*^{nuc} transgenic genetic labeling system, we produced single-cell transcriptomes from a mixed population of precursor cells at different developmental stages. We then developed a bioinformatic pipeline named Waterfall to reconstruct continuous biological processes at single-cell resolution using adult neurogenesis as our model, and applied this methodology to other stem cell datasets.

Chapter 2. Single cell RNA sequencing and Waterfall Analysis of Adult Neural Stem Cells[2]

[2] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

Single-cell RNA-seq of Neural Precursor Cells from the Adult Mouse Dentate Gyrus

In the adult dentate gyrus, radial glia-like qNSCs give rise to new neurons via a sequential process of activation, proliferation and generation of intermediate precursor cells (IPCs; Figure 2. 1A) (Ming and Song 2011). To elucidate detailed molecular dynamics during initial phases of adult neurogenesis in vivo, we used a transgenic mouse line that expresses nucleus-localized cyan fluorescent protein (CFP) under the Nestin promoter (*Nes-CFP*^{nuc}) (Encinas, Vaahtokari et al. 2006), which labels the majority of NSCs and their immediate progeny (collectively named NPCs; Figure 2. 2A). The SMART-seq protocol (Ramskold, Luo et al. 2012) was modified by adding DNase I treatment step to remove genomic DNA for single-cell cDNA amplification (Figure 2. 2A). In total, we performed single-cell RNA-seq for 142 CFP^{nuc+} and 26 CFP^{nuc-} single cells (Table 2. 1). Total RNA from wild-type adult mouse dentate gyri was serially diluted to 3 pg and processed in parallel for comparison.

We achieved, on average, 87% mapping onto annotated genes (Table 2. 1). Sequencing reads were evenly distributed throughout the whole span of transcripts with 3' bias comparable to recent studies (See Chapter 5. Single cell analysis with Waterfall). Correlation analyses of RNA samples from different batches indicated minimal technical fluctuation during cDNA amplification or across batches compared to significant biological heterogeneity among single-cell transcriptomes (Figure 2. 2B; Table 2. 1). Universally expressed genes, such as β-Actin/Actb, Gapdh, or Ubiquitin B/Ubb, showed even expression patterns across all individual cells, whereas known NSC markers Gfap and Sox2, or early IPC (eIPC) markers Tbr2/Eomes and Sox11, were expressed in subsets of cells (Figure 2. 1C; Table 2. 2).

Nes-CFP^{nuc} also labeled a small percentage of non-NPCs in the adult dentate gyrus (Figure 2. 2A and C). CFP transcript levels were multiple orders of magnitude higher in CFP^{nuc+} cells compared to CFP^{nuc-} cells or diluted dentate RNA (Figure 2. 1B). We excluded cells that exhibited markedly different transcriptomic profiles from the majority of the CFP^{nuc+} population or were clearly identifiable as non-NPCs, such as oligodendrocyte progenitor cells or pericytes (Figure 2. 2C). By differential expression analysis, we identified the top 35 genes enriched in CFP^{nuc+} cells, which included known NSC markers, Blbp, Spot14/Thrsp, Sox9 and GLAST/Slc1a3 (Table 2. 2 and Table 2. 3). In total, 31 out of 35 top genes exhibited SGZ-enriched expression patterns and/or

were known NPC genes (p = 6.1x 10^{-40}; hypergeometric test; Table 2. 3). These results provided initial validation of our approach.

Waterfall: Analyzing Single-Cell Datasets from Continuous in vivo Processes

We next examined the whole-transcriptome dataset of individual CFP^{nuc+} NPCs. Unsupervised hierarchical clustering analysis resulted in two super-groups with six sub-groups (Figure 2. 3A). Notably, these six CFP^{nuc+} groups were not clearly segregated on the principal component analysis (PCA) plot (Figure 2. 3B), which was consistent over different batches of sequencing runs with multiple biological replicates (Figure 2. 4A; Table 2. 1). The continuous trajectory called for a new approach not relying on segmentation into a few groups of cell clusters. We could not use currently available single-cell analysis software, such as Monocle (Trapnell, Williams et al. 2010) or Wanderlust (Bendall, Davis et al. 2014), for our system due to the lack of sufficient prior information, such as temporal delineators or a robust set of specific markers (See Chapter 5. Single cell analysis with Waterfall). We thus developed a more generally applicable pipeline of algorithms to perform unbiased statistical analyses of multidimensional single-cell datasets from continuous biological processes. We collectively named the suite of algorithms “Waterfall”, which involves three steps: pre-processing, pseudotime reconstruction, and gene expression analysis (Figure 2. 3C; See Chapter 5. Single cell analysis with Waterfall).

Pre-processing defined the trajectory of interest following dimensionality reduction of the data. Unsupervised learning identified six clusters of cells (Figure 2. 3A), which were then labeled S1-5 and SA based on their relative location in a PCA plot (Figure 2. 3B). SA was recognized as a branch by the minimum spanning tree (MST) algorithm (See Chapter 5. Single cell analysis with Waterfall). Although characterizing SA would be interesting (See Chapter 5. Single cell analysis with Waterfall), we focused on the major neurogenic pathway in the current study. The expression profiles of a few known developmental genes were used to orient the most probable trajectory of interest (Figure 2. 3B and Figure 2. 4B).

To reconstruct the chronology, we first determined the most probable route of transcriptomic progression. We performed k-means clustering of single-cell transcriptomes on the PCA plot after excluding SA, followed by constructing a MST trajectory to connect cluster centers (Figure 2. 6A). We then introduced “pseudotime” (Trapnell, Cacchiarelli et al. 2014) to define the relative location

of each cell on the MST trajectory (Figure 2. 6A). Taking the Euclidian distance defined by the whole transcriptomic difference from each cell to the next in pseudotime, we found that the total path length reconstructed by Waterfall was significantly shorter than would result from random ordering of cells (See Chapter 5. Single cell analysis with Waterfall). The pseudotime algorithm reconstructs molecular state transitions of a continuous process by quantifying the gradual divergence of single-cell transcriptomes individually, rather than as members of pre-classified groups.

For gene expression analysis, we developed an algorithm to determine the binary on/high or off/low expression state of each gene along pseudotime in an unbiased fashion using a hidden Markov model (HMM; Figure 2. 6B). Gene expression at the single-cell level is highly stochastic and binary (Novick and Weiner 1957, Raj, Peskin et al. 2006, Muramoto, Cannon et al. 2012). When analyzing continuous processes represented by single-cell transcriptomes in the absence of discrete groups, conventional statistical methods, such as arithmetic mean or t-test, are not appropriate. We adopted HMM to statistically convert stochastic expression patterns of individual genes into binary on/high or off/low states (Figure 2. 6B). The binary gene expression states were then shown as heat maps, which quantified the molecular cascade over time (Figure 2. 3C-D). To discover novel developmentally regulated genes, we correlated gene expression levels with pseudotime and subjected identified genes to gene ontology (GO) analyses (Figure 2. 3C).

Validation for the Reconstructed Adult Neurogenesis Process

We validated molecular dynamics revealed by Waterfall at multiple levels. First, known NSC markers Gfap and Apoe, and eIPC markers Sox11 and Tbr2, showed non-overlapping expression states over developmental pseudotime (Figure 2. 3D; Table 2. 2). Second, we evaluated in vivo expression of Aldoc and Stmn1, which have not been previously studied in adult hippocampal neurogenesis. Over pseudotime, Aldoc was initially highly expressed (on/high state), then downregulated (off/low state), whereas Stmn1 was initially off, then upregulated (Figure 2. 5A and C). In the hippocampal dentate subgranular zone (SGZ) of adult *Nes-GFP^{cyto}* mice (Encinas, Michurina et al. 2011), Aldoc^{+}GFP^{+} precursors were almost exclusively PCNA^{-} qNSCs, with very few PCNA^{+} active NSCs (aNSCs) or IPCs (Figure 2. 5B). In contrast, Stmn1^{+}GFP^{+} precursors were mostly eIPCs and PCNA^{+} aNSCs, but not PCNA^{-} qNSCs (Figure 2. 5D). Thus, Waterfall accurately

predicted the in vivo expression dynamics of both known and unknown genes. Third, for functional validation, we explored the possibility of genetic labeling of a specific developmental stage based on Waterfall results. Hopx (Hop homeobox) was highly expressed in qNSCs, but downregulated around the transition point from qNPC to aNSC (Figure 2. 5E). Upon a single low-dose tamoxifen injection into a *Hopx-CreER*^{T2} mouse line (Takeda, Jain et al. 2011) crossed with a *mT/mG*$^{f/f}$ reporter line for clonal lineage-tracing (Bonaguidi, Wheeler et al. 2011), almost all labeled precursors at three days post injection were nestin^{+}GFAP^{+} qNSCs in the adult SGZ (Figure 2. 5F). By 7 days, we were able to observe GFP-labeled clones that contained both NSCs and their progeny (Figure 2. 5G), indicating self-renewal and differentiation, two hallmarks of NSCs.

Table 2. 1. Summary of Sequencing statistics.

Batch names	Hiseq5	Hiseq6	Hiseq7/8*	Hiseq9	Hiseq10	Total
Run type	100 cycle single end	100 cycle single end	100 cycle single end	50 cycle single end	100 cycle single end	
Total read number	106,765,153	102,075,273	219,505,841	113,311,365	103,066,717	644,724,349
Number samples	19	24	53	34	42	172
Number CFP+ cells	19	19	39	30	35	142
Number CFP- cells	0	3	12	4	7	26
Number bulk RNA samples	0	2	2	0	0	4

Table 2. 2. List of literatures consistent with the predicted molecular dynamics from this study.

Developmental genes	
Spot14/Thrsp	Knobloch, M., Braun, S.M., Zurkirchen, L., von Schoultz, C., Zamboni, N., Arauzo-Bravo, M.J., Kovacs, W.J., Karalay, O., Suter, U., Machado, R.A., et al. (2013). Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis. Nature 493, 226-230.
Tbr2/Eomes	Hodge, R.D., Kowalczyk, T.D., Wolf, S.A., Encinas, J.M., Rippey, C., Enikolopov, G., Kempermann, G., and Hevner, R.F. (2008). Intermediate progenitors in adult hippocampal neurogenesis: Tbr2 expression and coordinate regulation of neuronal output. J Neurosci 28, 3707-3717.
Sox11	Haslinger, A., Schwarz, T.J., Covic, M., and Lie, D.C. (2009). Expression of Sox11 in adult neurogenic niches suggests a stage-specific role in adult neurogenesis. The European journal of neuroscience 29, 2103-2114.
Sox9	Scott, C.E., Wynn, S.L., Sesay, A., Cruz, C., Cheung, M., Gomez Gaviro, M.V., Booth, S., Gao, B., Cheah, K.S., Lovell-Badge, R., et al. (2010). SOX9 induces and maintains neural stem cells. Nature neuroscience 13, 1181-1189.
GLAST/Slc1a3	Doetsch, F. (2003). The glial identity of neural stem cells. Nature neuroscience 6, 1127-1134.
ApoE	Li, G., Bien-Ly, N., Andrews-Zwilling, Y., Xu, Q., Bernardo, A., Ring, K., Halabisky, B., Deng, C., Mahley, R.W., and Huang, Y. (2009). GABAergic interneuron dysfunction impairs hippocampal neurogenesis in adult apolipoprotein E4 knockin mice. Cell Stem Cell 5, 634-645.
Id3	Niola, F., Zhao, X., Singh, D., Castano, A., Sullivan, R., Lauria, M., Nam, H.S., Zhuang, Y., Benezra, R., Di Bernardo, D., et al. (2012). Id proteins synchronize stemness and anchorage to the niche of neural stem cells. Nature cell biology 14, 477-487.
Nr2e1/Tlx	Zhang, C.L., Zou, Y., He, W., Gage, F.H., and Evans, R.M. (2008). A role for adult TLX-positive neural stem cells in learning and behaviour. Nature 451, 1004-1007.
Hes1	Hatakeyama, J., Bessho, Y., Katoh, K., Ookawara, S., Fujioka, M., Guillemot, F., and Kageyama, R. (2004). Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. Development 131, 5539-5550.
SoxC (Sox4 and Sox11)	Mu, L., Berti, L., Masserdotti, G., Covic, M., Michaelidis, T.M., Doberauer, K., Merz, K., Rehfeld, F., Haslinger, A., Wegner, M., et al. (2012). SoxC transcription factors are required for neuronal differentiation in adult hippocampal neurogenesis. J Neurosci 32, 3067-3080.
Foxg1	Shen, L., Nam, H.S., Song, P., Moore, H., and Anderson, S.A. (2006). FoxG1 haploinsufficiency results in impaired neurogenesis in the postnatal hippocampus and contextual memory deficits. Hippocampus 16, 875-890.
Insm1	Duggan, A., Madathany, T., de Castro, S.C., Gerrelli, D., Guddati, K., and Garcia-Anoveros, J. (2008). Transient expression of the conserved zinc finger gene INSM1 in progenitors and nascent neurons throughout embryonic and adult neurogenesis. The Journal of comparative neurology 507, 1497-1520.
Insm1	Rosenbaum, J.N., Duggan, A., and Garcia-Anoveros, J. (2011). Insm1 promotes the transition of olfactory progenitors from apical and proliferative to basal, terminally dividing and neuronogenic. Neural development 6, 6.
Tcf12	Uittenbogaard, M., and Chiaramello, A. (2002). Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. Brain research Gene expression patterns 1, 115-121.
Nfib	Ninkovic, J., Steiner-Mezzadri, A., Jawerka, M., Akinci, U., Masserdotti, G., Petricca, S., Fischer, J., von Holst, A., Beckers, J., Lie, C.D., et al. (2013). The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. Cell Stem Cell 13, 403-418.
Dbx2	Pierani, A., Brenner-Morton, S., Chiang, C., and Jessell, T.M. (1999). A sonic hedgehog-independent, retinoid-activated pathway of neurogenesis in the ventral spinal cord. Cell 97, 903-915.
Nr3c1	Sundberg, M., Savola, S., Hienola, A., Korhonen, L., and Lindholm, D. (2006). Glucocorticoid hormones decrease proliferation of embryonic neural stem cells through ubiquitin-mediated degradation of cyclin D1. J Neurosci 26, 5402-5410.
Id4	Bedford, L., Walker, R., Kondo, T., van Cruchten, I., King, E.R., and Sablitzky, F. (2005). Id4 is required for the correct timing of neural differentiation. Developmental biology 280, 386-395.
Hmgb1	Merenmies, J., Pihlaskari, R., Laitinen, J., Wartiovaara, J., and Rauvala, H. (1991). 30-kDa heparin-binding protein of brain (amphoterin) involved in neurite outgrowth. Amino acid sequence and localization in the filopodia of the advancing plasma membrane. The Journal of biological chemistry 266, 16722-16729.
N-myc	Knoepfler, P.S., Cheng, P.F., and Eisenman, R.N. (2002). N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. Genes & development 16, 2699-2712.
Mxd3	Shimada, Y., Kuroyanagi, J., Zhang, B., Ariyoshi, M., Umemoto, N., Nishimura, Y., and Tanaka, T. (2014). Downregulation of Max dimerization protein 3 is involved in decreased visceral adipose tissue by inhibiting adipocyte differentiation in zebrafish and mice. International journal of obesity 38, 1053-1060.
Zeb2	Denecker, G., Vandamme, N., Akay, O., Koludrovic, D., Taminau, J., Lemeire, K., Gheldof, A., De Craene, B., Van Gele, M., Brochez, L., et al. (2014). Identification of a ZEB2-MITF-ZEB1 transcriptional network that controls melanogenesis and melanoma progression. Cell death and differentiation 21, 1250-1261.
ZT3/Zfp	Polimeni, M., Giorgi, S., De Gregorio, L., Dragani, T.A., Molinaro, M., Cossu, G., and Bouche, M. (1996). Differentiation dependent expression in muscle cells of ZT3, a novel zinc finger factor differentially expressed in embryonic and adult tissues. Mechanisms of development 54, 107-117.
Tsc22d3/Gilz	Bruscoli, S., Donato, V., Velardi, E., Di Sante, M., Migliorati, G., Donato, R., and Riccardi, C. (2010). Glucocorticoid-induced leucine zipper (GILZ) and long GILZ inhibit myogenic differentiation and mediate anti-myogenic effects of glucocorticoids. The Journal of biological chemistry 285, 10385-10396.
Hopx	Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. Science
Hopx	Takeda, N., Jain, R., Leboeuf, M.R., Padmanabhan, A., Wang, Q., Li, L., Lu, M.M., Millar, S.E., and Epstein, J.A. (2013). Hopx expression defines a subset of multipotent hair follicle stem cells and a progenitor population primed to give rise to K6+ niche cells. Development 140, 1655-1664.
Developmental signaling pathways	
Insulin/IGF2 signaling	Bracko, O., Singer, T., Aigner, S., Knobloch, M., Winner, B., Ray, J., Clemenson, G.D., Jr., Suh, H., Couillard-Despres, S., Aigner, L., et al. (2012). Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. J Neurosci 32, 3376-3387.
NMDA signaling	Cameron, H.A., McEwen, B.S., and Gould, E. (1995). Regulation of adult neurogenesis by excitatory input and NMDA receptor activation in the dentate gyrus. J Neurosci 15, 4687-4692.
Wnt signaling	Lie, D.C., Colamarino, S.A., Song, H.J., Desire, L., Mira, H., Consiglio, A., Lein, E.S., Jessberger, S., Lansford, H., Dearie, A.R., et al. (2005). Wnt signalling regulates adult hippocampal neurogenesis. Nature 437, 1370-1375.
BMP signaling	Lim, D.A., Tramontin, A.D., Trevejo, J.M., Herrera, D.G., Garcia-Verdugo, J.M., and Alvarez-Buylla, A. (2000). Noggin antagonizes BMP signaling to create a niche for adult neurogenesis. Neuron 28, 713-726.
Bdnf signaling	Pencea, V., Bingaman, K.D., Wiegand, S.J., and Luskin, M.B. (2001). Infusion of brain-derived neurotrophic factor into the lateral ventricle of the adult rat leads to new neurons in the parenchyma of the striatum, septum, thalamus, and hypothalamus. J Neurosci 21, 6706-6717.
GABA signaling	Song, J., Zhong, C., Bonaguidi, M.A., Sun, G.J., Hsu, D., Gu, Y., Meletis, K., Huang, Z.J., Ge, S., Enikolopov, G., et al. (2012). Neuronal circuitry mechanism regulating adult quiescent neural stem-cell fate decision. Nature 489, 150-154.
FGF signaling	Yoshimura, S., Takagi, Y., Harada, J., Teramoto, T., Thomas, S.S., Waeber, C., Bakowska, J.C., Breakefield, X.O., and Moskowitz, M.A. (2001). FGF-2 regulation of neurogenesis in adult hippocampus after brain injury. Proceedings of the National Academy of Sciences of the United States of America 98, 5874-5879.
Notch signaling	Pleasure, S.J., Collins, A.E., and Lowenstein, D.H. (2000). Unique expression patterns of cell fate molecules delineate sequential stages of dentate gyrus development. J Neurosci 20, 6095-6105.

Table 2. 3. Adult NPC-enriched gene list and validation based on the Allen Brain RNA in situ hybridization database.

Uncorrected p values (Wilcoxon rank sum	Corrected p values (Benjamini & Hochberg)	log2FD	Covar1	Covar2	TPM_CFP+, NPC	TPM_CFP-, neuronal	Gene_name	Allen atlas SGZ
3.12E-06	1.35E-03	5.513398304	0.866476274	2.05026401	593.1929231	12.98666667	Ptprz1	yes
8.45E-06	1.65E-03	7.148029451	1.032782959	1.697798523	6861.155154	48.37555556	Fabp7	yes
8.45E-06	1.65E-03	3.049352557	0.64142578	0.782681052	10472.36431	1265.022222	Dbi	yes
1.21E-05	1.73E-03	4.503347823	0.672894701	1.245446016	4355.766846	192.0533333	Slc1a3	yes
1.33E-05	1.73E-03	4.101642441	0.858969222	2.103493065	2185.322231	127.2911111	Cd9	yes
2.42E-05	2.90E-03	8.591404217	1.088297417	0.580234811	2893.316692	7.501111111	Thrsp	yes
4.11E-05	3.95E-03	4.349536335	0.802908031	1.792381448	1322.353615	64.86444444	Ppap2b	yes
9.08E-05	0.006641362	2.396843788	0.831080278	0.812084385	991.1899231	188.2066667	Hnrnpa2b1	yes
1.03E-04	6.76E-03	4.981848298	1.017756909	2.200124636	1228.779846	38.88555556	2610017I09Rik	yes
1.04E-04	0.006755994	3.205671563	0.862564264	1.840240225	1609.048538	174.4077778	Mmd2	ambiguous
1.09E-04	6.79E-03	2.795301582	0.82375671	1.506968619	658.4336923	94.85111111	Sox9	yes
2.06E-04	0.011947516	8.087390404	1.215233001	1.412191616	851.6194615	3.131111111	B2m	ambiguous
2.07E-04	0.011947516	2.150388132	0.737266688	1.170528541	2389.032846	538.1344444	Ndufc2	yes
2.31E-04	0.012839875	2.36910321	0.708327211	0.584880353	22575.10131	4369.77	Cst3	yes
2.56E-04	0.013731082	3.143811704	1.085206643	2.045264833	429.9273077	48.64222222	Hopx	yes
2.98E-04	0.013950151	3.996824191	0.912255588	1.516816198	4339.574538	271.8211111	Gstm1	ambiguous
2.98E-04	0.013950151	2.89511646	1.168269207	2.910818862	924.8958462	124.33	Cnn3	yes
3.11E-04	0.013950151	3.718420308	1.134249069	1.622670181	317.0097692	24.08333333	Dclk1	no
3.13E-04	0.013950151	9.318514394	1.669742053	1.559022993	439.1374615	0.687777778	Cdk4	yes
3.84E-04	0.016151265	9.906787182	2.415999817	2.471118293	117.3249231	0.122222222	2810055G20Rik	yes
4.01E-04	0.016430184	4.777984082	1.17486707	1.444870629	161.505	5.886666667	Pbx1	no
4.32E-04	0.016846943	5.77146291	1.811156967	2.116127144	109.9156154	2.012222222	Pitpnc1	no
4.33E-04	0.016846943	8.277186919	1.488233106	1.424655853	184.0688462	0.593333333	Eps15	yes
4.54E-04	0.016972126	2.772902073	0.802346253	1.126822052	836.5812308	122.4	Slc1a2	yes
4.58E-04	0.016972126	7.271714896	1.68785924	1.694498596	164.6570769	1.065555556	Pdpn	no
4.73E-04	0.017139429	4.043009581	1.058761678	2.774219046	1191.988846	72.31111111	Mfge8	yes
5.14E-04	0.017783341	2.309309182	0.868634517	1.133779149	1626.681615	328.1944444	Gpm6b	no
5.31E-04	0.017969691	7.76522177	1.4138105	1.804086119	367.6642308	1.69	Naaa	yes
6.15E-04	0.019550025	4.031442716	0.919928845	1.572202299	226.6098462	13.85777778	Gstm3	ambiguous
7.35E-04	0.022902946	7.315668711	1.124599768	1.012553256	1522.623923	9.557777778	Cmtm5	yes
7.66E-04	0.022912004	5.410329823	1.144439386	2.605218371	280.5408462	6.596666667	Ddah1	yes
7.79E-04	0.022912004	6.77820491	1.617221312	2.111178271	386.1103077	3.517777778	Id4	yes
8.07E-04	0.023273379	9.373295493	2.10550752	0.897732718	136.324	0.205555556	Tnc	yes
9.76E-04	0.027150686	1.762032045	0.705377953	0.763674623	4960.751	1462.591111	Ckb	yes

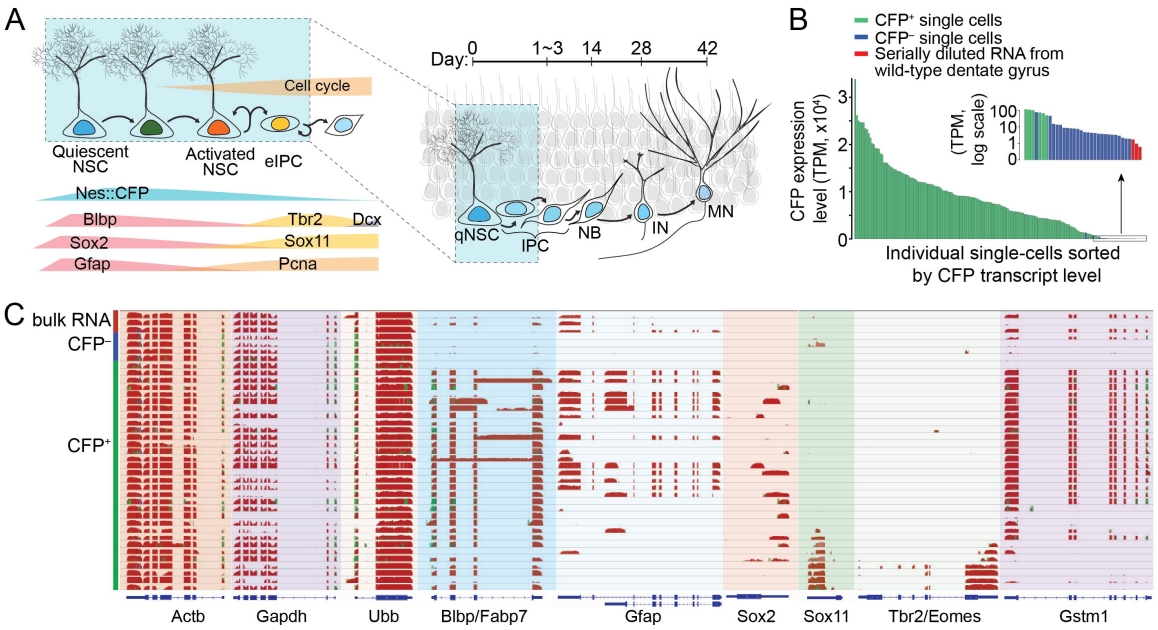

Figure 2. 1. Single-cell transcriptomes of adult neural stem cells and their immediate progeny.

(A) A schematic diagram of the process of adult neurogenesis in the dentate gyrus of the mouse hippocampus. Once quiescent neural stem cells (qNSCs) become activated (aNSCs), they enter cell cycle and generate early intermediate progenitor cells (eIPCs), which in turn give rise to neuroblasts (NB), immature neurons (IN) and then mature neurons (MN). Area highlighted with blue background indicates cell types fluorescently labeled in adult Nestin-CFPnuc animals.

(B) Expression levels of transcript encoding CFP in each single cell and diluted whole dentate RNA samples (TPM, transcripts per million). Inset: enlarged view of CFP transcript levels in logarithmic scale of samples with low abundance of CFP transcript.

(C) Representative coverage profile of diluted total RNA from the whole dentate gyrus, CFP– individual cells, and CFP+ individual cells at selected genomic loci, including house-keeping genes (β-actin/Actb, Gapdh, ubiquitin B/Ubb), known NSC markers (Blbp/Fabp7, Gfap, Sox2), known IPC markers (Sox11, Tbr2/Eomes), and potential new NPC markers (Gstm1).

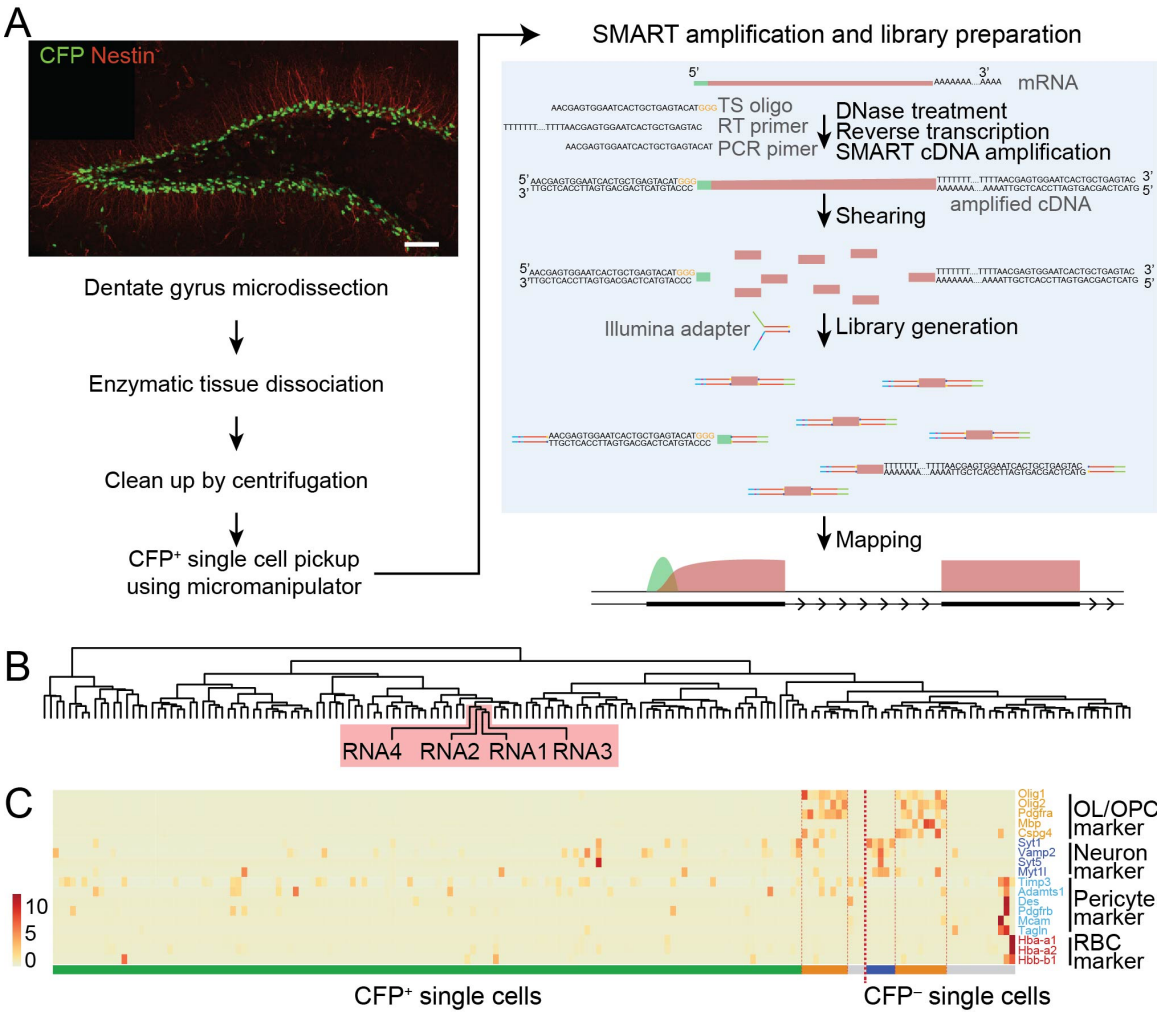

Figure 2. 2. Single-cell RNA-seq of labeled precursor cells from the adult mouse dentate gyrus.

(A) A schematic diagram of experimental procedures for achieving single-cell transcriptomes from fluorescently labeled individual cells isolated from the adult *Nestin-CFP*^{nuc} mouse dentate gyrus. Shown at the top is a sample confocal image of *Nestin-CFP*^{nuc} adult mouse dentate gyrus for CFP and Nestin immunostaining. Scale bar: 100 μm. Isolation of individual NPCs involved the following steps: (1) microdissection of tissues of interest; (2) dissociation of tissue using papain and DNase I; (3) eliminating cell debris via multiple rounds of mild centrifugation; and (4) picking up fluorescently labeled individual cells using a micromanipulator with a pulled glass micropipette and breaking the tip of the glass pipette into PCR strips. The SMART amplification protocol (Ramskold, Luo et al. 2012) was followed with minor modifications, including DNase I digestion before adding polyT primer. Library generation followed conventional Illumina library preparation after fragmentation of amplified cDNA. TS oligo: Template Switch oligo; RT primer: primer for reverse transcription; SMART: Switching Mechanism At 5' end of RNA Template.

(B) Unsupervised clustering for individual CFP^{+} and CFP^{-} cells, and single-cell equivalent amount (3 pg) of whole-dentate gyrus RNA samples (n = 4). Whole dentate RNA samples are highlighted to show the extensive biological variability of single-cell transcriptomes, compared to negligible technical variability.

(C) Expression heat map of known markers for oligodendrocyte precursor cells (OPC)/oligodendrocytes (OL), pericytes, and blood cells. Horizontal bar under the heat map represents putative cell types determined by marker expression profiles. Green: cells with potential NPC lineages; Orange: cells with OPC/OL lineages; Purple: cells with differentiated neuronal characteristics; Gray: uncharacterized outliers or cells with non-neuronal lineages.

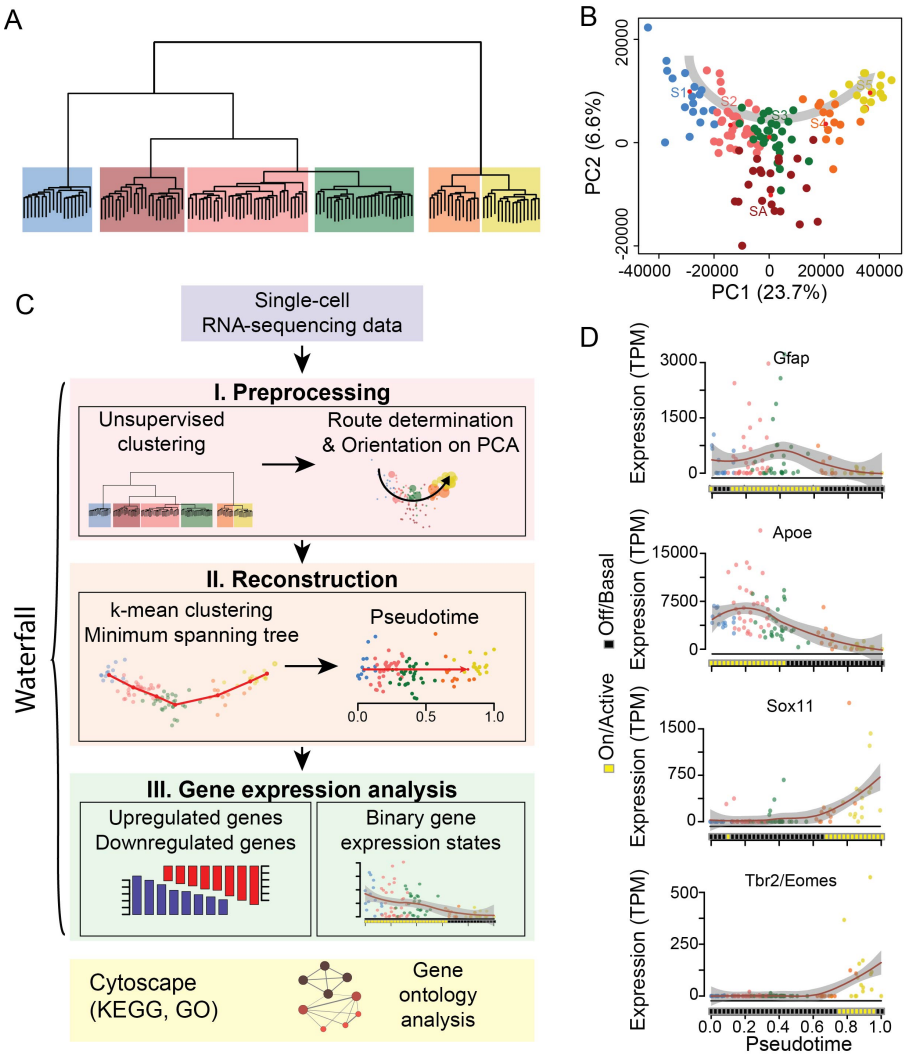

Figure 2. 3. Waterfall for analyzing single-cell data from continuous in vivo process.

(A) Unsupervised clustering analysis of CFP^{nuc+} NPCs resulting in two super-groups with six subgroups. Different groups are color coded in the same fashion in this figure and across all other figures.

(B) Principal component analysis (PCA) plot shows one of the possible linear trajectories of different groups with the exception of SA.

(C) A schematic diagram of multiple components and workflow of Waterfall. Waterfall is a full range of algorithms for processing multi-dimensional single-cell datasets derived from continuous biological processes. Please See Chapter 5. Single cell analysis with Waterfall for more information and Waterfall analyses of other biological systems.

(D) Representative expression profiles of marker genes of adult neurogenesis. Each data point represents the gene expression level of a single cell with color scheme following Figure 2. 3. Data points are fitted with local polynomial regression fitting (red lines) with 95% confidence interval (gray area). HMM-predicted underlying states are represented as black and yellow squares on the bottom of the graphs.

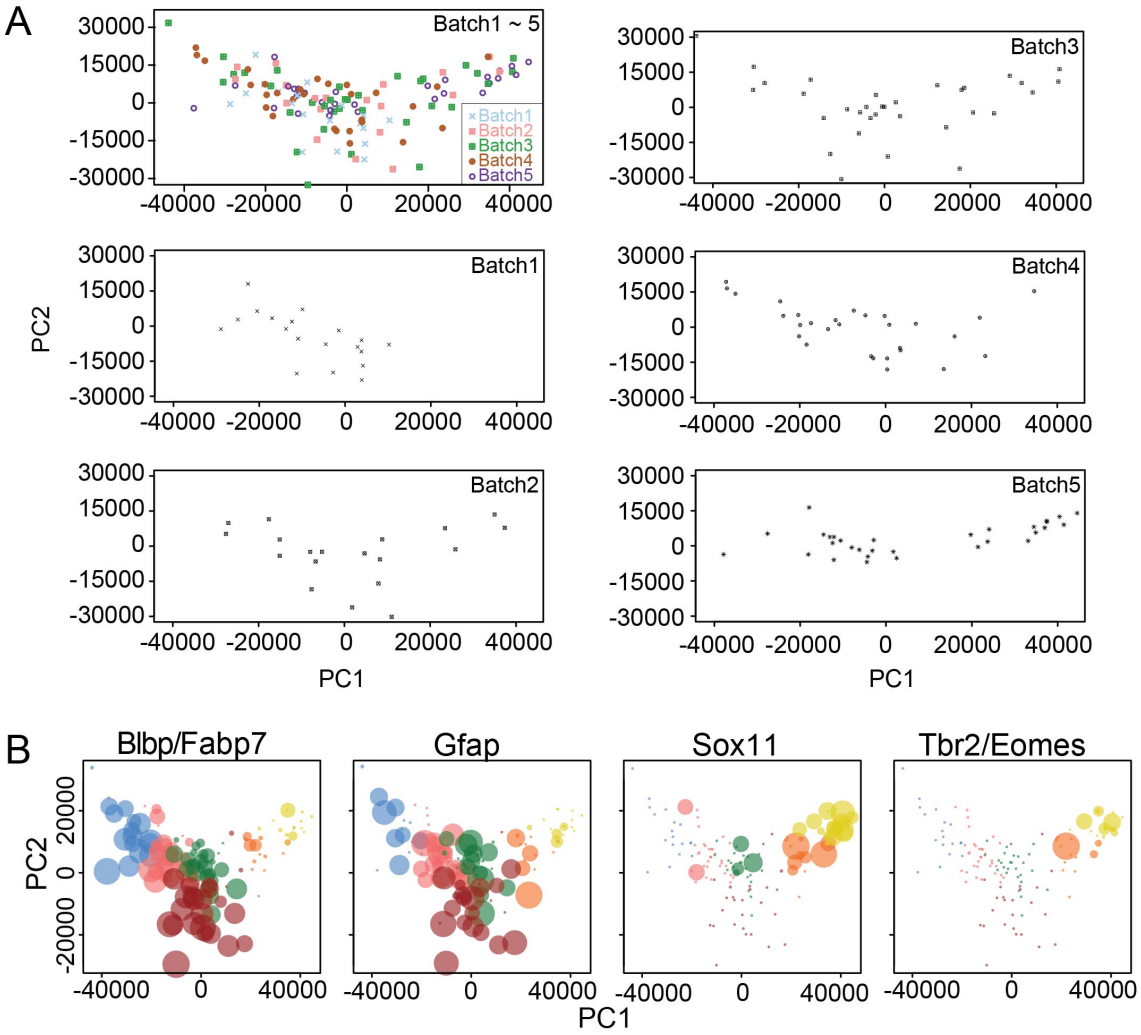

Figure 2. 4. Reproducibility and orientation of the developmental trajectory on the PCA plot

(A) Shown are PCA plots with all Nestin-CFP^{nuc+} NPCs (top left panel) and with Nestin-CFP^{nuc+} NPCs in five individual sequencing runs.

(B) Normalized expression levels of known marker genes on the PCA plot, represented by the size of data points. Colors of data points follow the color codes in Figure 2. 3. Notably, known NSC markers Blbp/Fabp7 and Gfap were highly expressed on the left side, whereas eIPC markers Tbr2/Eomes and Sox11 were highly expressed on the right side.

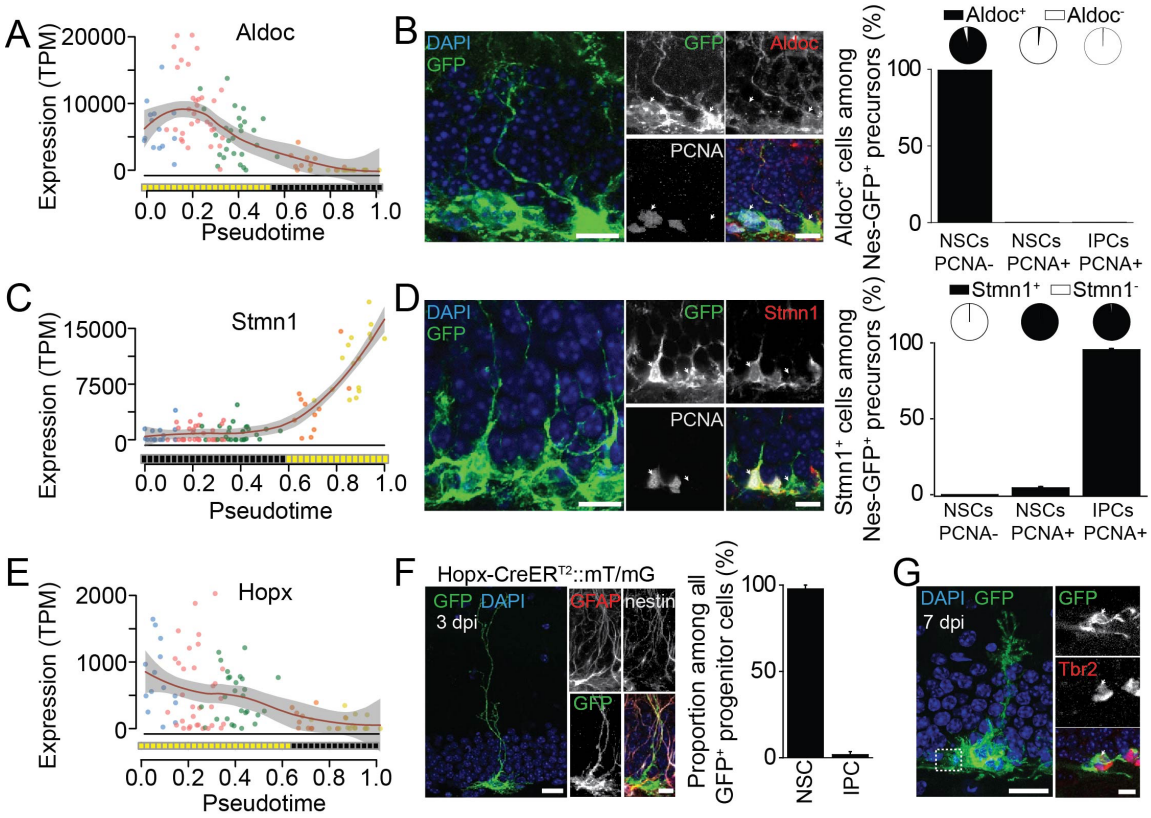

Figure 2. 5. Validation for Waterfall predictions for early adult neurogenesis.

(A-D) Validation of gene expression patterns and on/off binary states of Aldolase C (Aldoc, A) and Stmn1 (C) over pseudotime by immunohistology. Also shown are sample confocal images of GFP, cell proliferation marker PCNA, Aldoc or Stmn1 in the dentate gyrus of adult *Nestin-GFP^{cyco}* mice (left panels) and quantifications (right panels). Values represent mean ± SEM (n = 3 animals). The pie chart represents the proportion of Aldoc$^+$ or Stmn1$^+$ cells among each category of the GFP$^+$ progenitor population. Scale bars: 20 μm (left) and 10 μm (right).

(E-G) Validation of gene expression patterns and on/off binary states of Hopx over pseudotime (E) by genetic labeling and lineage-tracing. Adult *Hopx-CreER^{T2}::mT/mG* mice were injected with a single dose of tamoxifen and examined 3 (F) or 7 days (G) later. Shown in (F) are sample confocal images of GFP, GFAP, Nestin and DAPI. Also shown is quantification of percentages of GFP$^+$ cells as NSCs or IPCs. Values represent mean ± SEM (n = 5 dentate gyri). Shown in (G) is an example of a labeled clone containing an NSC and multiple Tbr2$^+$ neuronal progeny. Scale bars: 20 μm (left) and 10 μm (right).

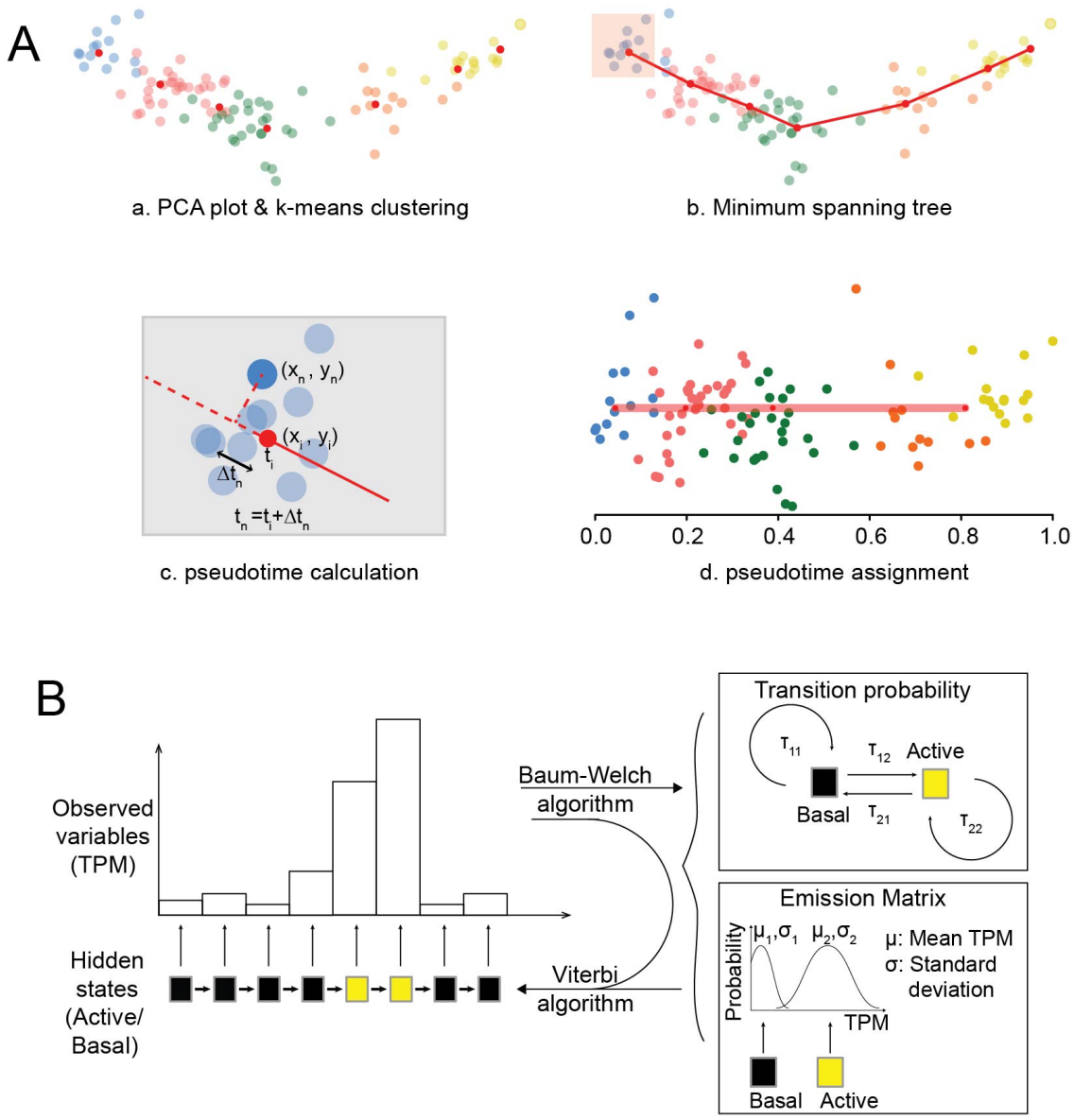

Figure 2. 6. Waterfall algorithms.

(A) An illustration of steps to generate an averaged trajectory and assign pseudotime to each individual cell: (a) Representation of individual cells on the PCA plot with PC1 and PC2 and performing k-means from the PCA data (small red dots); (b) Build a trajectory connecting k-means as minimum spanning trees (MST); (c) Determine the relative location of each cell using orthogonal line to connect each cell to the closest trajectory line; (d) Assign pseudotime for each cell based on its relative location on the trajectory.

(B) An illustration of the approach to predict underlying states from gene expression (TPM) over pseudotime progression. The Baum-Welch algorithm predicts the most likely transition probability and emission matrix from observed variables (TPM). The Viterbi algorithm uses observed variables (TPM) along with the output from the Baum-Welch algorithm to predict hidden On/High and Off/Low gene expression states.

Chapter 3. In Vivo Molecular Dynamics of Adult Neural Stem Cells.[3]

[3] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

Transcription Factor Expression during Adult NSC Activation and Neurogenesis

Waterfall allows an unbiased prediction of the relative chronological position of each individual cell and distribution of binary gene expression over the developmental trajectory. To delineate molecular cascades underlying adult qNSC activation and neurogenesis, we generated a list of the top 1000 negatively correlated genes with pseudotime ($DOWN^{1000}$ genes; Spearman correlation coefficient < –0.13; Table 3. 1A), which represent qNSC-enriched genes down-regulated during activation and neurogenesis, as well as the top 1000 positively correlated genes with pseudotime (UP^{1000} genes; Spearman correlation coefficient > 0.20; Table 3. 1B), which represent newly activated genes during qNSC activation and early neurogenesis.

Out of these 2000 genes, we initially focused on transcription factors (TFs). Most known TFs involved in adult neurogenesis were discovered by extrapolating findings from embryonic studies. In contrast, our database provides unbiased genome-wide profiles of TF expression. Systematic analyses of our dataset revealed a total of 41 down-regulated TFs and 42 up-regulated TFs during adult hippocampal neurogenesis (Figure 3. 1 and Figure 3. 2; Table 3. 2).

First, the set of dynamic TFs we identified included known regulators of adult NSCs and neurogenesis, which provided additional validation of our approach. Among DOWN TFs, Sox2, Sox9, Id3, nuclear receptor Nr2e1/Tlx, and Hes1 have been shown to regulate adult NSC maintenance and function. Among UP TFs, SoxC (Sox4 and Sox11), Foxg1, Tbr2, Insm1, Tcf12 and Nfib are critical in proliferative adult NPCs (Table 2. 2).

Second, we identified multiple dynamic TFs that are regulators of embryonic neurogenesis but have not yet been studied in adult neurogenesis. DOWN TFs include homeobox protein Dbx2, nuclear glucocorticoid receptor Nr3c1 and Id4. UP TFs included chromatin protein Hmgb1 and proto-oncogene N-myc. During embryonic neurogenesis, these DOWN TFs inhibit cell cycle (Nr3c1) or prevent premature differentiation (Id4), whereas UP TFs regulate progenitor proliferation (Hmgb1, N-myc), suggesting conserved functions during embryonic and adult neurogenesis (Table 2. 2).

Third, more than half of UP TFs and DOWN TFs are largely uncharacterized in the context of neurogenesis, but many of them are close paralogs or binding partners to other neurogenesis-related genes, or have been implicated in other somatic stem cell systems. Examples include

Hmgb1 paralogs (Hmgb2, Hmgb3, Hmga1-rs1), SWI/SNF-related Brg1/Smarca4-associated factors (Smarcc1/Baf155, Smarce1/Baf57), and Nfib paralogs (Nfia, Nfix). In addition, Mxd3, Zeb2 and ZT3/Zfp, regulate adipocyte, melanocyte and myogenic differentiation, respectively, whereas Tsc22d3/Gilz inhibits myogenic differentiation (Table 2. 2). Hopx, which we found to mark qNSCs in the adult dentate gyrus (Figure 2. 5E-G), is expressed in intestinal stem cells and a subset of multipotent hair follicle stem cells (Table 2. 2).

Together, our single-cell transcriptome datasets are a rich resource of genome-wide dynamic expression profiles of TFs during adult neurogenesis. Many TFs that were previously uncharacterized in adult neurogenesis are known to be involved in embryonic neurogenesis or regulation of other somatic stem cells, suggesting shared biology among different stem cell systems and the potential utility of our resource for the general stem cell field.

Molecular Cascades underlying Adult qNSC Activation and Neurogenesis Initiation

The vast majority of UP^{1000} and $DOWN^{1000}$ genes in our dataset were not TFs (Table 3. 1). We investigated their characteristics from three perspectives: transition patterns along the developmental trajectory, cellular location of gene products, and biological function.

For transition patterns, plots of the top 150 genes each from UP^{1000} and $DOWN^{1000}$ lists showed a wave of molecular activation or inactivation events over time, highlighting the sequential transition of gene expression during qNSC activation and neurogenesis (Figure 3. 1B). To obtain biological insight into these transition patterns, we performed multiple gene ontology analyses. Strikingly, the predicted cellular localizations of protein products of UP^{1000} genes and $DOWN^{1000}$ genes were drastically different. 51% of $DOWN^{1000}$ genes, as opposed to 20% of UP^{1000} genes, encode proteins associated with the membrane (Figure 3. 3A, p value = 2.6×10^{-29}). On the other hand, 58% of UP^{1000} genes, as opposed to 20% of $DOWN^{1000}$ genes, encode proteins associated with the nucleus (Figure 3. 3A, p value = 1.2×10^{-36}). Similar results were obtained using different thresholds for generating lists of UP and DOWN genes (Figure 3. 4A).

Molecular Signatures of Adult qNSCs Revealed by DOWN^{1000} Genes

Functional annotation of DOWN^{1000} membrane genes revealed enrichment for ion or protein transport, cell communication and cell adhesion (Figure 3. 3B). Further classification identified proteins specific to the plasma membrane, endoplasmic reticulum, Golgi apparatus, and cytoplasmic vesicles (Figure 3. 4B). KEGG pathway analysis revealed diverse functional entities involved in intra- and inter-cellular communication (Figure 3. 3B). Specifically, Notch signaling, GABAergic synapses, glutamatergic synapses, BMP pathways, MAPK pathway, calcium and cell adhesion-related genes were down-regulated upon qNSC exit from quiescence (Figure 3. 3B, Figure 3. 5A-B and Figure 3. 6A). Electrophysiological recordings of *Nestin-GFP*$^{cyto+}$ NSCs in acute slices from adult animals showed responses to both AMPA and NMDA, suggesting expression of functional receptors (Figure 3. 5B). Each functional signaling pathway entity contained key genes that encode receptors, subunits or downstream mediators (Figure 3. 5A-B). Notably, many ligands for these receptors, including glutamate, GABA, Wnts, BDNF/neurotrophin, Jagged1, BMPs, FGFs and Insulin/IGF2, are known to be present in the adult SGZ niche, suggesting active signaling in qNSCs (Table 2. 2). While previous studies have examined each of these ligands and receptors in regulation of adult neurogenesis in isolation, our systematic genome-wide analyses unified disparate information and suggested a novel model that quiescent adult NSCs are not passive or dormant, but instead are actively integrating various niche signals. More surprisingly, qNSC activation was associated with decreased expression of genes involved in transducing local environment cues and pervasive down-regulation of various signaling pathway-related genes (Figure 3. 5A). These results suggest that, once activated, adult NSCs shunt their capacity to respond to external regulation.

KEGG analysis of DOWN^{1000} genes also revealed a shift in energy source and metabolism. First, multiple lipid metabolism-related functional entities, including fatty acid degradation and sphingolipid metabolism, were enriched in qNSCs, but down-regulated upon activation (Figure 3. 3B). As previously reported (Knobloch, Braun et al. 2013), qNSCs exhibited the highest level of Spot14 (Figure 3. 6B), which regulates lipid metabolism. qNSCs also maintained an active fatty acid degradation pathway (Acsl3, Acsl6, and Acsbg1; Figure 3. 6B; Table 2. 2). Second, pathway analysis consistently indicated glutathione metabolism and glycolysis as an adult qNSC characteristic, which was lost upon activation. Among glycolysis genes, aldolase A, aldolase C, and

Ldhb decreased significantly, whereas most other glycolysis genes including Gapdh did not change during initiation of neurogenesis (Figure 3. 6C; Table 2. 2).

To validate results from analyses of the top 1000 significantly down-regulated genes, which contained a limited number of genes in each particular pathway, we performed analysis using all expressed genes and an independent functional annotation database wikipathway (Pico, Kelder et al. 2008). Virtually identical results were obtained (Figure 3. 8). Together, analyses of down-regulated genes provided novel insight into molecular signatures of adult qNSCs, including both intrinsic properties and regulation of intra- or inter-cellular signaling pathways.

Sequential Molecular Dynamics during Adult Neurogenesis Revealed by UP^{1000} Genes

We next analyzed UP^{1000} genes, which were nucleus-associated and/or related to cell cycle, DNA/RNA metabolism and chromosome organization (Figure 3. 3A). Detailed analysis revealed pervasive activation of cell cycle-related genes, ranging from cell cycle supporting genes, such as nucleotide synthesis, protein/RNA synthesis, and DNA fidelity controls (DNA repair and p53 signaling pathways), to genes directly involved in cell cycle, such as DNA replication, kinetochore complex, cyclin/cyclin-dependent kinases, or cytosolic mitotic spindle genes (Figure 3. 3B).

As opposed to down-regulation of glycolysis-related genes, oxidative phosphorylation-related genes were up-regulated (Figure 3. 6C). Specifically, in contrast to stable expression of earlier mitochondrial respiratory chain complexes (complex I, II, III and IV), expression of subsequent complexes (complex V) increased over pseudotime, implying a gradual completion of the full electron transport chain during neurogenesis (Figure 3. 6C).

The high resolution of Waterfall analyses revealed temporal relationships among genes in different functional groups. Cell cycle checkpoint genes were sequentially activated following the known biological sequence of cell cycles: G1 to S transition, followed by G2 to M transition and then chromosomal segregation, indicating that our pseudotime accurately reconstructs sequential biological events (Figure 3. 6D). Initiation of cell cycle preceded the major transcriptomic shift (Figure 3. 5C). Notably, up-regulation of genes encoding ribosomal subunits preceded the appearance of any cell cycle checkpoint genes (Figure 3. 5C and Figure 3. 6D), suggesting that

priming of protein synthesis machinery may mark the G0 to G1 transition ahead of cell cycle entry during adult qNSC activation.

Together, analyses of UP^{1000} genes suggested that molecular dynamics of qNSC activation and initiation of neurogenesis are largely defined by priming of protein synthesis machinery, cell cycle entry, activation of RNA and protein biogenesis, and a shift in energy metabolism from glycolysis to oxidative phosphorylation. An independent approach using a different functional annotation database showed similar results (Figure 3. 8).

Holistic Picture of Molecular Cascades underlying Adult Neurogenesis Initiation

Based on the molecular dynamics from qNSCs to aNSCs and then eIPCs, we have reconstructed sequential waves of biological events from single-cell RNA-seq data and Waterfall (Figure 3. 7). The process begins with adult qNSCs down-regulating transcription factors defining quiescence and decreasing competence for cell signaling (RTKs, GPCRs, neurotransmitter receptors, cytokines, calcium). Concurrently, glycolysis, glutathione and fatty acid metabolism begins to wane, while up-regulation of protein translation capacity is the first marker of a pre-activation stage. As NSCs enter cell cycle, oxidative phosphorylation becomes the primary energy source. Progression through cell cycle accompanies a major decline in NSC metabolism (glutathione, fatty acid, drug metabolism) and an increase in eIPC transcription factors. Finally, kinetochore and chromosomal segregation occurs in the first neurogenic progeny. Overall, the developmental trajectory is defined by a coordinated switch from a membrane-targeted to a nuclear-targeted transcriptome, suggesting a transition from qNSCs dominated by extrinsic signaling to eIPCs dominated by a pre-programmed intrinsic molecular cascade.

Table 3. 1. 1000 UP genes and 1000 DOWN genes and their Spearman correlation coefficient to pseudotime.

Upregulated genes

Hmgn2	0.698025803	Gins2	0.427797125	Fern1	0.373586281	Cmc1	0.329581851	Sox4	0.300924826	Nras	0.275910902	Vbp1	0.258944905	Shmt2	0.242540962	Ankrd46	0.226392249	Umps	0.21974487
Tmsb4x	0.663746287	Ckslb	0.424246594	Smc6	0.373544081	Arl2bp	0.329401448	Serbp1	0.300908854	9430037G07Rik	0.275510392	Snx22	0.258936681	Snrnl	0.242529979	Zwilch	0.226396985	Arpc2	0.21793823
Ptma	0.64152452	Slc24a5	0.423952036	Mcm5	0.373373901	Bub1b	0.329351537	Ndel	0.300372385	4930503E24Rik	0.275405103	Acot13	0.258906058	Pxk2l	0.242437331	Ncpb1	0.226219069	Rtc2	0.21785987
Tmsb10	0.62838941	Ppia	0.42370237	Bex2	0.373365169	Anln	0.329188821	Qser1	0.299819867	2700099C18Rik	0.275296055	Ppp3ca	0.258637296	Psma2	0.242403367	Tsen34	0.226229216	Fili	0.21768703
Stmn1	0.61124525	Rpl26	0.423950227	Suma2	0.372870473	Lsm2	0.32902342	Ptbp4	0.299785953	Smarca5	0.275226965	Rbm41	0.25737823	Nup107	0.24210156	Fkbp1a	0.225769176	Ppih1	0.21507515
Eef1a1	0.607944794	Tuba4a	0.423122851	Top1mt	0.372804693	Ppp1r14b	0.32890957	Tef1	0.299655162	Satb3	0.274860811	Rpl28	0.257178449	Rwdfl	0.241828106	Fkbp2	0.225656983	Lsm6n	0.21501804
H2afz	0.59619118	Rpl27	0.421688133	Cfdp1	0.372211411	Hexl	0.328846333	2700029M09Rik	0.29939041	Nsa2	0.274761635	Gdi2	0.257047442	Bub3	0.241826776	Anapc4	0.225654994	Ints10	0.21491137
1190002F15Rik	0.594837701	Hnrnpab	0.420225159	Haus1	0.37143492	Mrpl13	0.328214257	Aqpn	0.298093543	Ccq95	0.274185833	Ifitm2	0.256884156	Trp53	0.240756899	Tmem57	0.225210158	Cnp1	0.21483041
Rps24	0.570074868	Rps14	0.419314963	Ddx39b	0.371344366	Ccdc18	0.327898801	Rpcd1	0.298079699	BC056474	0.274159107	Cyp39a1	0.256744457	Tmem254c	0.240688457	Rtn4	0.225288655	Mrps28	0.21395646
Rpl32	0.56660959	Nfix	0.419222499	Syncrip	0.370907767	Ube2s	0.327388448	Pcin2k	0.297953109	Ddx1	0.274061605	Prickle3	0.256731213	Nup205	0.240454858	Mrpl14	0.225072943	Dynll1	0.21123123
Rpsl9	0.56539991	Rps7	0.417354778	Dpy30	0.369396285	Pola1	0.327242743	Mxi1	0.297340686	Spg11	0.27398376	Dvl1	0.256591274	Rpl10l	0.239988895	Tcof1	0.225010023	Rcm42	0.21097496
Hmgb2	0.563914521	Rpl19	0.415945271	Haus4	0.368150505	Rpl24	0.327092662	Rpl34-ps1	0.297057894	Mettl9	0.273811919	Cdc42	0.256532423	Abl1	0.239855612	Dbn1	0.224755848	Ccp110	0.21088415
Hmgb1	0.561644164	Rpl6	0.415866063	Nap1l1	0.368011905	Mtfp4	0.326971511	Gdpd1	0.296824414	Mycn	0.273390111	Zmiz1	0.256422275	Psmd10	0.239588046	Cnot6	0.224714079	Smarca4	0.21085916
Rpl15	0.55336392	Ezh2	0.415412975	Alyref2	0.367205423	Rdx	0.326845317	Exosc9	0.296577764	Hnrnpl2	0.27337269	Hspb9ad	0.256369637	Smc3	0.239537298	Pcnp	0.224695184	Dnajb9	0.21077694
Mki67	0.553097237	Arp32b	0.415258304	Gnb2l1	0.366485069	Rpl34	0.326411583	Rpl36a	0.296076048	Rabl2	0.27302828	Depdc1a	0.256340427	Taftd	0.239449146	Dcakd	0.224482758	Pcbp2	0.21035666
Rrm2	0.548347025	2810025M15Rik	0.414484254	Rps15a-ps4	0.366459646	Mrl18a	0.325997555	Cenpf	0.295812522	Tsn	0.273025398	Kars	0.255619812	Snrpc	0.239402351	Nudt5	0.224462692	Alas1	0.21023929
Ccnd2	0.544362436	Rpl37	0.412644594	Fam69a	0.366381966	Tcf19	0.325900478	Tmem254c	0.295655951	Dnajc8	0.272775871	Hmgn5	0.255602291	Trpc6	0.239229626	Kxd1	0.224373625	Unk	0.21006909
Rpl41	0.538093068	Fkbp3	0.412000849	Psmc7	0.366320657	Cdk2ap1	0.325672608	Tdp1	0.295138455	Fnrl1	0.272645501	Lsm4	0.255487347	Ndufa5	0.23873612	Ddx1	0.224253671	Anp32e	0.21000619
Cdca8	0.534650343	Ier5	0.411202695	Nasp	0.365625902	Ube2d2a	0.32540434	Kpna3	0.294585826	Rpl11	0.272631259	Rbmx	0.255376716	Ube2c3	0.238583733	Cdipt	0.224002698	Uba2	0.20996615
Tubalb	0.52788637	Kif23	0.410995576	Elavl2	0.36568179	Odc1	0.325320188	R3hdm1	0.294525964	Cdkn2c	0.272540128	Akr1a1	0.255238686	Rtl5	0.238485522	Btc1	0.224398836	Ciao1	0.20985856
Rpl35	0.516259238	Dtl	0.410743443	Smarcc1	0.365256938	Shfm1	0.324743566	Txndc16	0.293582565	Rufy3	0.272479903	Rab11b	0.255210881	Vtlb	0.238474537	Emc4	0.223834481	Rap1b	0.20971711
Dctpp1	0.514775606	Rnaseh2c	0.408741704	Rrp9	0.364011784	Rab10	0.323852585	Ges5	0.293184481	Rfk	0.272455629	Fam103a1	0.255187646	Yrb6	0.238374453	Actb	0.22370237	Rpa3	0.20968350
Lzic7	0.513836162	Dynll1a	0.408616596	Mrpl12	0.363494353	Tra2b	0.323634171	Amez2	0.29302408	Vps28	0.272418587	Nudt2	0.254921594	Pxk3k	0.238351023	Dcunl1l	0.223618584	Anapc2	0.20965570
Snrpg	0.513258883	Gpd1	0.407956056	Smc4	0.362818234	Gtk6	0.323465496	Tcf4	0.292889289	Lsm12	0.272011688	Slbp	0.254568254	Igsf8	0.238143432	Cxcr4	0.223484825	Nuf2	0.20959594
Tubb5	0.509428471	Cenpa	0.407479957	Tmem55b	0.36244882	Rpsl3	0.32273075	Mrpl34	0.292476073	Tcf20	0.271948398	Echs1	0.254440548	Ddx50	0.238105136	Kpna2	0.223455975	Elf2a	0.20944859
Tpm4	0.508496737	Fnbp1l	0.40647712	Anapc13	0.36151553	Med3	0.322617394	Tuba4	0.292351514	Wdr35	0.271936959	Ginsl	0.254375288	Tspo	0.238090744	Acin1	0.223324721	Ccdc124	0.20925761
Myef2	0.506961002	Snhg5	0.406035939	Arpc5	0.36135793	Pbdc1	0.322550513	Akip1	0.292221508	Atp5a1	0.271803996	Sephs2	0.254266435	Slval	0.238039325	Rbm10	0.222922315	Apfg1	0.20924666
Hist1h2ao	0.505927028	Cenpn	0.404872995	Hnrnpa3	0.36122327	Rpl10a	0.322445129	Eif3g	0.291851924	Pdhx	0.271589957	AY358078	0.253867158	Nsmce1	0.23782978	Hspb11	0.222905787	Aimp1	0.20905046
Sox11	0.502698374	Snrpe	0.40441318	Ddx39	0.361220825	1700123O20Rik	0.321884493	Igfbp11	0.291774695	Fzvg1	0.271499678	Eef1g	0.253724002	Pacs	0.237538146	Pxk2m	0.22251522	Snrnp25	0.20902933
Uba52	0.501435998	Snrpdl	0.403936542	Snrpb2	0.360948173	Zbed3	0.3217281	Marcksl1	0.291454856	Lima1	0.271408449	Rps27	0.253676978	Tmem161b	0.237476432	Pnn	0.222210803	Dntttip2	0.20899175
Dut	0.50119526	Rps3a1	0.403642749	Cdca3	0.360908274	Commd1	0.321588721	Wtap	0.291308785	Ginl3	0.271107586	Ivns1abp	0.253655503	Smg6	0.237206202	Fam162a	0.222110839	Gpcpd1	0.20879435
Rpl37a	0.500438124	Rpl18a	0.403601571	Pdia6	0.360877993	Sumo3	0.321420348	Emd	0.291099489	Chchd6	0.270977807	Wdr70	0.253593	Mkrn1	0.236951843	Mrpl49	0.222069067	Cmtm2a	0.20876982
Ybx1	0.499839469	Naa38	0.403441438	H3f3b	0.360590223	Cdkn3	0.321399715	Sdhc	0.290771172	Mtch2	0.270701837	Tardbp	0.253397366	Snrpd3	0.236856671	Klhl13	0.221761056	Rpl7a	0.20864399
Rps11	0.496567282	Serf1	0.401986421	Anapc15	0.360064113	Exosc8	0.321319759	Tpm3	0.290491108	Sync2	0.270422959	Fhl4	0.253007681	Bird8	0.236772388	1110004F10Rik	0.221759712	Stom2	0.20862979
Rpl13a	0.496033603	Elavl4	0.401202459	Hsfsp3	0.359966257	Zfp62	0.320913646	Psph8	0.290368764	Naa10	0.270414791	Ccdc34	0.252600165	St3gal1	0.236659442	Rpl27a	0.221759145	Srhq12	0.20857194
Esco2	0.49559327	Nusap1	0.400747769	Rad51ap1	0.359903568	Eif2s2	0.32083812	Psmd2	0.290050537	Txnl4a	0.270231834	Atp5b	0.252584016	Prmt5	0.236621507	Elf1a	0.221674748	Rrm4	0.20852051
Cdk4	0.492750996	Cenpk	0.400506364	Nup54	0.359898856	Hnrnph1	0.320168017	Rpl30	0.290000817	Hspa14	0.270203281	Lenc2	0.252416485	Mina	0.236620986	Pole3	0.221628356	Cbx3	0.20840467
Rbm3	0.490797839	Pcna	0.400128631	Tex30	0.359845206	Rbm8a	0.319166149	Aass	0.289798494	Cdc20	0.269888462	Nfyc	0.252381316	2410006H16Rik	0.236193132	Bora	0.221428349	Asnsd1	0.20827452
Top2a	0.490729057	Snrpb	0.400050664	Rpsl0	0.359643415	Rnaseh2b	0.319511548	Cbx1	0.289727123	Pds5b	0.269741334	Elf3l	0.252043248	Def8	0.236054421	Fus	0.221242875	Sarla	0.20810461
Rpsl2	0.47846535	Cbx5	0.399726367	Ncaph	0.35942204	Nono	0.31917496	Tcp1	0.289659736	Eif3l	0.269663015	Usp10	0.251636266	Lars2	0.235957719	Ubqln2	0.220939556	Polr2i	0.20810775
Rpl3	0.47719719	Mecks	0.399504347	Ncapd2	0.357621315	Ncapg2	0.318778938	Akr1b3	0.288880559	Dnajc2	0.269639147	Zfp386	0.251468817	Psmc1	0.235934669	Iqw	0.221009412	Cmsl1	0.20780875
2810417H13Rik	0.476955184	Hist1h1b	0.398998195	Kif4	0.357160274	Ctnnb1	0.318710313	Hsd17b7	0.288552301	Klf15	0.269511111	Psma5	0.251419083	Sep3	0.235858044	Kdm2b	0.220737041	1110038F14Rik	0.20771225
Rfc4	0.476592627	Tubb5	0.398775878	Rps8	0.356748936	Ptpl2	0.318573775	Psmd12	0.288369106	Cdca7	0.269455234	Prc1	0.251345135	Npm1	0.235368252	Cep19	0.220653292	Mcef6	0.20750882
Nop58	0.476332127	Insm1	0.39876028	Blf3	0.356612228	Flgn1	0.317468052	Ncaph2	0.288092423	Atp5g2	0.269376162	Actl6a	0.251252948	Wdr26	0.235306812	Cdkn2aipnl	0.220579617	Tbca	0.20745661
Alyref	0.475686867	Ccnb1	0.39759261	Shcbp1	0.356589719	Ranpkp1	0.317058495	Ganab	0.288032764	Mcm3	0.269007518	Fbxo5	0.251134884	Aafla	0.235245588	Pwp1	0.220446436	Ppp2cb	0.20663915
Rps5	0.475609503	Tirrc7	0.397008594	Eiflax	0.356057264	Gatm2	0.316899021	Syn2	0.287850077	Arhsa2	0.268814547	Hspa5	0.250986311	Eif4a3	0.235210338	Mrps10	0.220336981	Sgta	0.20657715
Rps26	0.475415154	Dnmt1	0.396364491	Dym	0.355009555	Tpx2	0.316847434	Cenpc1	0.287834907	Cenpn	0.268807401	Rbm18	0.250769484	Sass6	0.235176475	Rhed3	0.220249943	Tmed4	0.20653616
Mtap2	0.47512381	Ints7	0.395659127	Zfp367	0.354978058	Psrp1	0.316590695	Tmt1	0.287573327	Ppm1g	0.268697167	Pdxhlb3	0.25068343	Nap1l4	0.235059145	Ccng2	0.220248834	Rhno1	0.20651427
Rps20	0.474682206	Bc2l12	0.395221244	Ywhaq	0.354502768	Tmpo	0.316461004	Wdr67	0.287277433	Impa4	0.26855176	Ubb	0.250453045	Fbxl2	0.234936404	Sod2	0.220060925	Mapk7	0.20649808
Lsm5	0.474095059	Uhrf1	0.395185033	Nudcd1	0.354330873	Spag5	0.316321013	Hdgf	0.287173763	Xpo1	0.268468363	Gas2l3	0.250394542	Setd8	0.234759245	Nosip	0.220032722	Copé6	0.20639078
Tpt1	0.472923669	Rpl9	0.395052326	Prdx2	0.353876621	Cdk1	0.315967168	Pstpl	0.287156716	Naf1	0.268382161	Twsg1	0.250339599	Sec61g	0.234756003	Zfp808	0.219583913	1110001A16Rik	0.20622790
Mcm7	0.472437302	Tmsx	0.39488868	Trim59	0.353695052	Rpa4	0.315934559	Psmd7	0.286770564	Csl	0.268276585	Zkscan7	0.250231874	Ndufs6	0.234488959	Samd8	0.219568616	Orc3	0.20610731
Pcbp1	0.472422545	C330027C09Rik	0.394183868	Cisd2	0.353514039	Uchl1	0.315593766	Trim28	0.286218758	Usp1	0.268216508	Bzw2	0.250076273	Trappc2	0.234417753	Zbtb38	0.219448142	Gpsm2	0.20590033
Cct7	0.470400087	Ned3	0.393412768	Racgap1	0.352282736	Top1	0.315463063	Dtx9	0.286142246	Swi5	0.268024849	Nedd4	0.250021002	Txndl	0.234263145	Hirit	0.219119822	Nsmce2	0.20577229
Rps9	0.467883325	1810009A15Rik	0.393233649	Tacc3	0.353639643	Pxdc2	0.315166517	Cep170	0.285853114	Chchd1	0.267624364	Wdr76	0.249677539	2700049A03Rik	0.234119515	Rnxkb	0.219026878	Mcd3	0.20568778
Ranbp1	0.465681933	Ncl	0.392974476	Timm50	0.351591585	Dnajc9	0.314734323	Hspa4	0.285801429	Kif22	0.267476527	Rpp21	0.249656158	Snf9	0.233674811	Psmb7	0.218888331	Bnip3l	0.20558022
Nmral1	0.465640602	Klhl7	0.392217977	Eeftb2	0.351170966	Mnos1	0.314351108	Tubgl	0.285788467	Ngfrap1	0.267424501	Snx1	0.249609218	Aplm1	0.233389088	Tbp1	0.218843174	Mtx2	0.20543697
Srt	0.46529853	Hmgd	0.392144282	Rpl22l1	0.351052768	Hspb2ab1	0.314084351	Sox2	0.285131929	Eiflb	0.267283597	Rps4x	0.24933115	Psmd1	0.233204149	Trappc9b	0.218369499	Pdrl	0.20536009
Wdr89	0.463937786	Comm3	0.391002547	Rtc3	0.349058657	Srsf3	0.312790663	Cdca4	0.284508386	Kpnb1	0.267173327	Mrps11	0.249295366	Ankrd40	0.232941653	Atf4	0.218108517	Hdgfrp2	0.20518743
Ckap2l	0.462992843	Rps15	0.390543684	5830418K08Rik	0.348771658	Mepre1	0.312371923	Mstol	0.284316238	Psmc3ip	0.266976234	Ankmy2	0.248397886	2610316D01Rik	0.232789305	Mrpl36	0.217995665	Casp2	0.20494958
Smc2	0.46008355	Mad2l1	0.389900957	Actg1	0.348538854	Fbl	0.311561833	Nusl1	0.284099695	Fxr1as1	0.266240198	Bcl7a	0.248225041	Cdin2d	0.232563106	Naca	0.217908679	Timm10	0.20491854
Incenp	0.459446755	Sf3b2	0.388753697	Lyar	0.348335785	Acp1	0.311180446	Mdh2	0.284038299	Ncapd3	0.265962911	Snf10	0.248161556	Sep11	0.232509401	Mif	0.217954881	Nudt21	0.20480812
Rplp2-ps1	0.459345935	Ran	0.388710005	Pin1-ps1	0.347878948	Rps25	0.310969756	U2af2	0.283885949	Bclaf1	0.265839101	Tdp2	0.24811557	Cuedc2	0.232384759	Rbm12b2	0.217714331	Nedd1	0.20479029
Eef2	0.458951074	Hist1h1e	0.388520562	Nme1	0.347382287	Rpl5	0.310357658	Magoh	0.283864386	Cdc25c	0.265361992	Mrps18c	0.248067722	Hyou1	0.2320166	Dhx32	0.217570567	Znhit6	0.20428455
Hnrnpa1	0.458944964	Vrk1	0.38845704	Iqgap1	0.347085431	Psma1	0.309802785	Rpl12	0.283859799	Haus3	0.265000102	Spc24	0.247920804	Ndufc1	0.231980274	Ipo5	0.21763269	Fbxo42	0.20407662
Fam64a	0.45860673	Rps15a-ps6	0.388050021	Rpl13	0.345590706	Tyms-ps	0.309418182	Pcf11	0.283659862	Cdc42se2	0.264989891	Znrd1	0.247800137	Eftud2	0.231875519	Nsfl	0.217180612	Psma4	0.20392702
Rpl22	0.458033016	Pcbd4	0.38786279	Aurka	0.345107323	Dhx15	0.309369363	Eif3d	0.283473601	Sars	0.264969728	Nubp1	0.247659355	Tomm22	0.231850813	Nmt1a	0.217180367	Tars	0.20387973
Rps3	0.457281728	Ipo11	0.387842604	Dpysl3	0.34415466	Rps2	0.309247293	Srp19	0.283332527	Lonp1	0.264898293	Mphosph8	0.24766754	Cenpv	0.231842851	Vrk3	0.216873757	Brl3	0.20376478
Tubb3	0.456080693	Calm2	0.387374359	Orc6	0.343269067	Zeb2	0.308995735	Arhgef2	0.283322983	Tcf12	0.264704018	Bccip	0.247497764	Csde1	0.231818223	Bag1	0.216778279	H2afy	0.20357832
E030024N20Rik	0.454377438	Hist1h2ag	0.38678491	S100a11	0.342939796	Gtf2a2	0.308371277	3110039M20Rik	0.283085676	Tnfaip1	0.2644218	Myo5a	0.247453444	Ef3a	0.23170271	Ddx52	0.216449626	Hsph1	0.20349848
Rpl14	0.453086746	Eif4h	0.385601049	Suptl6	0.34206717	Banf1	0.308134326	Nbn	0.283039808	Mtl13	0.264012711	2610301B20Rik	0.247282871	Clf2	0.231592552	Leo1	0.216347459	Dnajb6	0.20348643
Rrm1	0.452482399	Ddx21	0.385576679	Cascs	0.340230529	3010026O09Rik	0.307846366	Encl	0.282621723	Dynlll1	0.263786539	Supt5	0.247001219	Acaf1	0.231525958	Serf3	0.216276564	Anapc5	0.20343084
Rpsa	0.452325233	Rbbp4	0.385492918	Dek	0.34078171	Bzw1	0.307392823	Rpl10	0.281589549	Lage3	0.263580379	Eif3h	0.24684012	1190038B21Rik	0.231079882	Myh10	0.216101145	Dnajc5	0.20340751
Spc25	0.451621665	Cnc2	0.384860836	Cdk2	0.340065826	Hat1	0.30718484	Kdm3b	0.281480456	Tmem177	0.263439265	Rbmx2	0.246718652	Ddb1	0.230850612	Man1	0.216088779	0610009B22Rik	0.20339692
Hells	0.45408907	Lrrcc1	0.384596294	Gadd45g	0.339932679	Katnal1	0.306997107	Eif3a	0.28025413	Mpp6	0.263365906	Tceal	0.246588749	Zfp422	0.230547655	Hmmph3	0.215870183	Sms	0.20326403
Ccna2	0.450786592	Hmgb3	0.38444457	1700037H04Rik	0.33988755	Iif2	0.306228648	Hmgn4-rs1	0.281166079	Celsg	0.263092768	Tyms	0.246420642	Zfp326	0.230460013	Tomm34	0.215850553	Sim	0.20296521
Rps6	0.449192919	Snhg6	0.383881794	Psrc1	0.339238396	Rps2f	0.306912018	Sugp1	0.280925306	Rpl7	0.262841438	Sun1	0.246361011	Ppib	0.230458845	Ndufs7	0.215772761	Hdac2	0.20281926
Cspl	0.447821756	Snrpd2	0.383482421	Rpsl8	0.339012848	Cks2	0.306899517	Zfp637	0.280863146	Sae1	0.262741751	Rnf114	0.246199567	Ccdc47	0.229840236	Slain2	0.215574135	Ural	0.20279578
Ckap2	0.447547645	Sgol1	0.383310557	Syca2	0.338905825	Rpa1	0.305630735	Znhit1	0.280839814	Arpp19	0.262615523	Tmem38a	0.245894782	Arlpap2	0.229730475	Osl4	0.215485855	Phb2	0.20278454
Cenpw	0.446592243	Rpl4	0.382622262	Rpl35a	0.338834917	Mrpl47	0.305624578	Elfa	0.280788267	Nhp2	0.262566634	Zfp763	0.245677417	Mdk	0.229697957	Tceal	0.215205365	Snrnp200	0.20272486
Pebp1	0.446353346	Fxyd6	0.382511571	Ndufs6	0.337856617	Lsm3	0.305584775	Rpo27a	0.280721441	Bax1b	0.262320645	Abcf2	0.245593724	1810022K09Rik	0.229607795	U2af1	0.215062168	Bhmt	0.20271827
Rplp2	0.444401623	Apex1	0.381562854	Cdca2	0.337779525	Mcm4	0.305530373	Gtf2h2	0.280176968	Zfp955a	0.262235165	Rpl36	0.245486289	Oser1	0.229474361	Ict1	0.214700276	Ppat	0.20249628
Dtymk	0.444428902	Rps8	0.381274707	Smarce1	0.337729006	Khdrbs1	0.305233194	Nudt14	0.280634888	Arrdc3	0.262080208	Pgdpf	0.245397256	Rcc2	0.229182982	Trim44	0.214648575	Rbl	0.20209774
Rplp0	0.440854634	Adss	0.380990328	Ap5	0.337498442	Tk1	0.305055782	Gpn1	0.280467991	Snhg1	0.261907469	Hmmpu	0.245055782	Ube2l	0.229110151	Nudt53	0.214334502	Sdhd	0.20196845
Rpl23	0.44069607	Ylhdf2	0.379681572	Rplp1	0.337272394	Ptms	0.304547628	Hnrnpm	0.280278993	Slc35a1	0.261802376	Srp9	0.245028973	Ppp1r7	0.228969901	Slc35c2	0.214056491	Copg1	0.20191633
Rpl23a	0.440001855	Rad51	0.379234289	Kmt5a	0.336989298	Tlc3	0.304426096	Thumpd1	0.280201236	Gltbp2	0.261786116	Ncoa4	0.24485869	Cdk5rap3	0.22889163	Cmpk1	0.213963215	Cenpb	0.20179975
Fmnl4b	0.438458933	Mis18bp1	0.378485112	Eif4el	0.336853159	Nfia	0.304409441	Kdm1b	0.279984054	Trim35	0.261610687	Cenpq	0.244689174	Ece2	0.228748624	Fcf1	0.213672324	Ebna1bp2	0.20167697
Nfib	0.438033117	Cenpq	0.378393944	Armc10	0.335956886	G3bp1	0.304319908	Sst8	0.279493736	Kifl1	0.261354176	Lsm7	0.244232687	Glod4	0.228548283	Psmb2	0.213657214	Fyttd1	0.20149553
Coro1c	0.437417569	Ssrp1	0.378197982	Hmgn1	0.335561901	H3f3a	0.303810381	Nabp2	0.279406886	Pgpic3	0.261140898	Tmem97	0.243942379	Gspt1	0.228506649	Pox3	0.213600071	Cse1l	0.20148928
Ptk6	0.436523991	Hnrl	0.377727954	Hmgxb	0.335245424	Coa3	0.303793945	Cald1	0.279274223	Cdsp2	0.261080856	Zfp964	0.243742036	Ccdc174	0.228413678	Pnisp3	0.213438363	Mpdl6	0.20138163
Rpl39	0.436160908	Snrnp27	0.376891128	Vps37b	0.335096992	Btg2	0.303157993	Cpix2	0.278696218	Emg1	0.261071096	Tctex1d2	0.243501693	Uqcrh	0.227924024	Coq7	0.212779552	Setd5	0.20129363
Etl	0.434735543	Ndrg3	0.376057622	Smim4	0.334104524	Hnrnpf	0.303000937	Snrpf	0.278687479	Cul2	0.26070061	Zbtb41	0.243457817	Usp37	0.227843761	Psmb3	0.212706907	Grcc10	0.20128972
Atpif1	0.433081463	Cox7c	0.376002349	Slpl	0.332977066	Ccs8	0.302610345	Tmem160	0.278430436	Birc5	0.260688493	Bud31	0.243479747	Phf1	0.227810669	Coa5b	0.212689696	Txnp	0.20118018
Hist1h1c	0.431909759	2700094K13Rik	0.375702284	Rpl31	0.332950294	Rpl21	0.302311138	Mtpn	0.278350761	Dck	0.26067829	6720489N17Rik	0.243413634	Hnrnpr	0.22747124	Bloc1s2a	0.212608755	Palp2	0.20115118
Rps16	0.431430355	Elf3f	0.375534866	H2afy	0.332651261	Cycs	0.301801991	Eny2	0.278103596	Hnrpdl	0.260411121	Senp8	0.243133885	Ehmt1	0.227345025	2410015M20Rik	0.212502176	Gpaa1	0.20102801
Lgl1	0.430665567	Pfdn4	0.375331742	Ube2l3	0.332000935	Brf1b4	0.301661544	Psmc6	0.277712641	Hpltp3	0.260309228	Atox1	0.243150933	Sv6	0.226928813	Slc7a5	0.212423576	0610010K14Rik	0.20099431
Rps17	0.430415463	Rad23a	0.375136215	Gtf2h5	0.331478297	Txn1	0.301604005	R3hdm2	0.277457239	Golt1b	0.260087086	Tceg1	0.242973959	Zmiz2	0.226868579	Slc35b2	0.212368413	Srt	0.20065224
Cenpp	0.429677512	Mpl30	0.374366211	Wee1	0.33077341	Mrpl3	0.301465925	Smc1a	0.276960105	Eomes	0.260055393	Lin9	0.242895658	Pebp1	0.226655215	Sugp2	0.212290778	Rbm14	0.20060449
Brd3	0.4290106	Pund14	0.374178994	Ptf6	0.330605775	Rnpc1	0.300251663	Strbp	0.276775454	Cog2	0.259620511	Ube3a	0.242872234	Lin7c	0.226634242	Marf	0.212195669	Map2	0.20050832
Rpl31-ps12	0.428305027	Mcm6	0.374051774	Hist1h2ac	0.330014724	Mrpl28	0.301207062	Polr2j	0.276102011	Dynll1b	0.259330847	Arfl	0.242631626	Ppox	0.226469579	Park7	0.212064908	Nsg2	0.20053899

Downregulated genes

Cst3	-0.971173117	Prex2	-0.382586259	Glul	-0.296752247	Arhgef26	-0.257391688	Garem	-0.22442398	Mmd	-0.20099906	Taf9b	-0.184689589	Uggt2	-0.168721239	Enol	-0.154746029	Tmem144	-0.142383671
Tspan7	-0.768200821	Arhgap5	-0.379056454	Ntm	-0.295519906	Gjc3	-0.256491188	Msrb2	-0.224498929	Acadm	-0.200228305	Amotl1	-0.184670326	Xirx	-0.168575363	BC018242	-0.154673266	Tmcc3	-0.142369362
Sic1a2	-0.736117612	Epn15	-0.378777675	Rtn4r20	-0.294603465	Atp2c1	-0.256292968	Tpm4b	-0.223896496	D330041H03Rik	-0.200067551	Prpsap1	-0.184590283	Usp33	-0.168563046	Slc7a10	-0.154652934	Tmem126a	-0.14227351
Slc1a3	-0.719219922	A630089N07Rik	-0.377893789	Acadl	-0.293618976	Cxcl14	-0.255722362	Rab34	-0.223885408	Bckdhb	-0.199706641	Tmem135	-0.184394419	Rapfa	-0.168530294	6720456H20Rik	-0.154092173	Abat	-0.141566536
Ppap2b	-0.710411041	Dkk3	-0.375700959	Glrx2	-0.293453345	Pde8b	-0.255538926	Ssu72	-0.223875434	Naa16	-0.19926491	Ikzf2	-0.184074076	Slc22a4	-0.16846771	Impact	-0.154024334	Psrp4	-0.141496152
Gstm1	-0.706558656	Timp2	-0.375325196	Mras	-0.293235289	Mcph1	-0.255515732	Ngef	-0.22379486	4833427F06Rik	-0.199533068	Tbc1d	-0.183977966	Pxwr	-0.168455067	Gsta4	-0.153948177	Avil	-0.141421763
Gstm3	-0.705769699	Icosl	-0.374008164	Acsbg1	-0.292600419	Tomm11	-0.255134717	Cyp2d22	-0.223725013	Fam63b	-0.199145035	Usp8	-0.183916819	Dcaf17	-0.168407851	Cntnap3	-0.153657903	Prima2	-0.141386105
Atp1a2	-0.69713774	Acot11	-0.373543788	D830030K20Rik	-0.291758458	D4Wsu53e	-0.254887689	Lpar1	-0.223441344	2300009A05Rik	-0.199142827	Smim8	-0.183873077	Gnh1	-0.168329305	Unc119	-0.153609961	Left	-0.14122781
Tlyn1	-0.69446953	Nkain4	-0.37243099	Mr1	-0.289503616	A730077B15Rik	-0.25471471	Ednrb	-0.223085847	Prdx1	-0.19879988	Tpst2	-0.183729104	Cwf19l	-0.168101781	Usp21	-0.153307895	Flswt1	-0.14113473
Aldoc	-0.687212721	Lmbrd1	-0.368919746	Ocrl	-0.288708135	St6galnac5	-0.253793709	Herpud2	-0.223082469	Tmed7	-0.19874603	Trim2	-0.183438894	Tk2	-0.167874762	Alkbh7	-0.153074432	Fam49a	-0.140936616
Apoa1	-0.675975598	Cd302	-0.368856863	Bmyc	-0.288539375	Selfl	-0.253667124	P4ha1	-0.222770734	Arsb	-0.19863975	Ubxn4	-0.183339674	Accs1	-0.167733466	Slc35a4	-0.153062607	Apobec3	-0.140796776
Clu	-0.66038009	Jagl	-0.368529427	Mlc3	-0.288003409	A730017L22Rik	-0.253297133	Prpsap1	-0.222759525	Art3b	-0.19847747	Camk2a	-0.182814564	Dstyk	-0.167669178	Slc7a10	-0.152906337	Rin2	-0.140504798
Itm2b	-0.659717972	Lsamp	-0.367655364	Kdm3a	-0.287768943	4933402H24Rik	-0.253016176	Kiflb	-0.222496379	Slc30a1	-0.19830078	Nfx1	-0.18276651	Bbip1	-0.167597263	Scara3	-0.152859107	Mertk	-0.140395233
Mslat1	-0.642472247	Sdc4	-0.367378391	Prss23	-0.287693443	E330013P04Rik	-0.252985394	Slitrk2	-0.222347438	Ecl1	-0.19778035	Hsdl1l1	-0.182585315	1700029J07Rik	-0.167496279	Tox4	-0.152808531	Stard13	-0.140313978
Naa9	-0.639453984	Dclk1	-0.366648665	Hist1h2bc	-0.287513736	Ai314180	-0.252850992	Fgfr1	-0.222321742	Ptprs1	-0.19778964	2900052N01Rik	-0.182539805	Lonrf3	-0.167481875	Etv5	-0.15273238	Rab7	-0.140216021
Ptprz1	-0.635943594	Cpe	-0.36631261	Gatm	-0.287370094	Mcc	-0.252155754	Per3	-0.222232983	Npdc1	-0.197532683	Trip11	-0.182401143	Sh2b1	-0.167440289	Chd2	-0.152593594	Cd320	-0.140178112
Ntrk2	-0.608616862	Dtna	-0.365296913	Gaa	-0.285999867	Sfxn5	-0.251750839	Rasgrp1	-0.222201589	Pik3ip1	-0.196796659	Cyp20a1	-0.182262588	Maip4	-0.167264512	Aar2	-0.152402242	Rtm4b	-0.140125127
Gas1	-0.590909954	Lgr4	-0.364166248	Emc1	-0.285864131	Tril	-0.25104361	Gpc1	-0.221870229	9530026F06Rik	-0.19639824	Epc2	-0.18272386	Necap2	-0.16751889	Tmcc1	-0.152155914	Hba-a1	-0.140104648
Fabp7	-0.585442544	Nwd1	-0.364157146	4930470H14Rik	-0.285436544	Hmp19	-0.250915327	Tspan18	-0.2218352	Ppm2	-0.19632452	Ptpns3	-0.182088593	Bmpr1b	-0.16711465	Nbp7	-0.151923192	4933427D14Rik	-0.139857585
Car2	-0.582628736	Btg1	-0.363901482	Rc3h2	-0.28534986	Prkcdbp	-0.250883565	Asb3	-0.220992423	Tlr3	-0.196258086	Supt7l	-0.181863146	Sgce	-0.166925984	Dagl	-0.151826579	Grin2b	-0.139676535
Plp2g7	-0.580457787	Satl	-0.362571807	Slc35e2	-0.285076494	Slc38a6	-0.250523614	4930579G24Rik	-0.220540363	Fam108b	-0.19566250	Cxcr2	-0.18173473	Pold3gl	-0.166890446	Herc3	-0.151783032	Golph3l	-0.13964413
Gpm6b	-0.576693669	Cntnap2	-0.361557092	Itm2c	-0.284759692	Cabiea1	-0.249981337	Katnal1	-0.220283946	Tusc3	-0.19559289	Usp2	-0.18163823	Lcmt1	-0.166239377	Pik2	-0.151678153	SemaSb	-0.139571328
Ptn	-0.564421522	Scd2	-0.360020102	Lxn	-0.284394682	Cmlf	-0.249904807	Ctsf	-0.22007893	Aplp2	-0.19538925	Zmynd8	-0.181422142	Gas1	-0.166180051	Mier1	-0.151654496	Dvl1	-0.139484002
Dbi	-0.55919592	Slc4a1	-0.359135548	Pfkfb2	-0.283510127	Tnfrsf10b	-0.248660966	Rcn2	-0.219976531	Rnf182	-0.19538019	Kctd	-0.181387628	Dctn1c	-0.16615114	Rnf149	-0.151625545	Thap7	-0.139364595
Gk	-0.556987699	Ckb1-10	-0.355600088	Famfl8a	-0.282760565	Stxbp3a	-0.24872761	Tbcel	-0.219895892	Akt2	-0.195351642	Atp6v0c	-0.181256301	Ifngr1	-0.166072779	Akap8	-0.151580144	Trim36	-0.139355133
Sparcl1	-0.555787579	Flsd3	-0.355285354	D630033O11Rik	-0.282537853	Ppp1r12c	-0.248234463	Rcbg1	-0.219689623	Dagla1	-0.19533254	Ndufc2	-0.180900009	Idua	-0.165809882	Fut2	-0.151538641	Tlr5	-0.13932444
Hepacam	-0.545009927	Ctnnd2	-0.355954742	Cpne2	-0.282292932	Bst1	-0.247380224	S100a13	-0.219670703	0610031J06Rik	-0.195250001	Klhl24	-0.180938236	Nhs	-0.165500812	Scamp1	-0.151513191	Tmc6a	-0.139209562
Cd81	-0.539525953	C030018K13Rik	-0.354917039	Tmem50a	-0.282220704	Necc2	-0.246800756	Unc13b	-0.21963846	Lipi	-0.195161808	Alg6	-0.180739487	Txndc15	-0.165050297	2010015L04Rik	-0.151471027	Tyw5	-0.13912966
Slka1	-0.533970999	Arglu1	-0.354743474	Ifi8	-0.281321457	Armc1	-0.246432486	Fbxl5	-0.219605265	Gfap	-0.19512437	Ech1d2	-0.180665382	Armh3	-0.164840278	Tmem245	-0.151448496	Fstl5	-0.138983866
Scrg1	-0.528780566	Lcat	-0.353629243	Cdh11	-0.280976328	Sbds	-0.246080694	Elmod3	-0.219569259	Cbr1	-0.194937524	Pla2g2c	-0.180656882	Adcy1	-0.164687752	Trat1	-0.151268968	Mcf1	-0.138967886
Mlc1	-0.522507524	Pcdh10	-0.352504779	Pdpn	-0.28059791	Dnajc6	-0.246069914	Slc23a2	-0.219287502	Kbtbd11	-0.194669897	Cox4i1	-0.180437457	2610005L07Rik	-0.164644241	Acot2	-0.15120612	Stam	-0.138874003
Gpm6a	-0.514515452	Spry2	-0.352192954	Ccdc30	-0.280410121	Dgkg1	-0.244667364	Scd3	-0.218624357	Reln30	-0.194614818	Dnc2	-0.180273141	Dcaf13	-0.164524118	Scp2	-0.15114018	Snx32	-0.138863706
Id4	-0.51041423	Ndrg2	-0.351829294	Ifit3	-0.279747786	Gabbr1	-0.244465336	Ncam2	-0.218230724	Wdr85	-0.194606044	Usp35	-0.180002113	Klf15	-0.164226623	Klc1	-0.151005145	2810407C02Rik	-0.138794876
Cspg5	-0.505061877	Scg3	-0.351263126	Acadvl	-0.279713061	Tcta	-0.243655385	Ylpf4	-0.21822459	Pmm1	-0.194558867	Tnfrsf12	-0.17999466	2210039B01Rik	-0.164019328	Pscd4	-0.150991171	Asah1	-0.138709689
Mt3	-0.502772444	Ascl1	-0.348024045	Ptpladl1	-0.279453339	Abca8b	-0.243653551	Tpsnp	-0.217910679	Bcl6	-0.194367483	Tmem251	-0.179610165	Ndufa12	-0.163925772	Jakmip1	-0.150938937	Atp1b1	-0.138668216
Sparc	-0.49710171	Tmem176b	-0.346706756	Mccc2	-0.278815793	Zfp277	-0.243427073	Ifnar1	-0.216537654	Sorbs3	-0.194365266	Vps13b	-0.179520756	Pxdc1	-0.163865695	Pxdc1	-0.150878471	Pxdc1	-0.138668216
Prdx6	-0.494797748	Ddah1	-0.34551867	Armcx3	-0.278614502	Rfx4	-0.243128808	Gpc6	-0.21625039	Tcf25	-0.194360922	Vps13b	-0.179488111	Zbtb26	-0.163647078	Dscam1	-0.150449831	Lrp6	-0.138099801
Atp1b2	-0.49211692	Cav2	-0.346972671	Mageh1	-0.278582895	Tspan9	-0.242697634	Dbx2	-0.215893784	Tst	-0.19413499	Ccdc92	-0.179367756	1110046J04Rik	-0.163623997	Epha5	-0.150433828	Syn3	-0.13901788
Gnao1	-0.489997392	Appl2	-0.343802061	Cdh1	-0.278385551	Mtor1d2	-0.242577167	Psap	-0.214914011	Crot	-0.194004	Icpsc1	-0.179331041	Gskip	-0.163456144	4933408N05Rik	-0.150294972	Tbc1d14	-0.138009301
Prnp	-0.48870087	Camk2d	-0.342852673	Fmn2	-0.277329189	Stxn4	-0.242312402	Pcyt2	-0.214892338	Casc4	-0.193975398	Vmn1r65	-0.179339684	Lpp	-0.163451816	Gbp3	-0.150192437	2810474O19Rik	-0.137995188
Stpr1	-0.48517561	Cbs	-0.342427325	Dadl1	-0.276452674	Echdc1	-0.242257602	Ranbp3l	-0.214802479	Magt1	-0.193579358	2900097C07Rik	-0.179091583	Ovol3	-0.163257408	Gnaq	-0.149984994	Syngr1	-0.137963707
Chst2	-0.484173321	Fam107a	-0.341238243	Kctd4	-0.276414487	Decr2	-0.241851087	Dnajc18	-0.214723158	Stox2	-0.193567416	Mfap3l	-0.178681158	Tm9sf4	-0.163022619	Gpt2	-0.14979118	Cpped1	-0.137445665
Notch2	-0.483342134	Aqp4	-0.339947334	Ctdspl	-0.276187	Asrgl1	-0.241620463	Klhl5	-0.214596832	Tspan3	-0.193489929	Dennd2a	-0.178553485	4932438A13Rik	-0.162965969	Prorsd1	-0.149614548	Spred1	-0.137362116
Hopx	-0.482357178	Ppargc1a	-0.339382573	Prodh	-0.2761808	Rab1	-0.241369306	Sycp2	-0.214507842	Ofcc1	-0.193403434	Calm1	-0.178145815	Akap6	-0.16294276	Srrm2	-0.149589897	Fam120c	-0.136904795
Luzp2	-0.480368008	Grk4	-0.339525953	Znrf3	-0.275541122	D130020L05Rik	-0.24115624	Ralgapb	-0.214398999	Map6	-0.193361	Ppp1r3c	-0.177995948	Cln3	-0.162805869	Gusb	-0.149563618	Ncoa7	-0.13656031
Ckb	-0.475007501	Creg1	-0.338903035	Bcan	-0.274214229	Cadm2	-0.240814329	Flad1	-0.21412627	Cbx3	-0.193321297	Tysj23b	-0.177976154	Rabgap1l	-0.162325378	Copa7a	-0.149547007	Mxi1	-0.136447878
Timp3	-0.468502811	Nme7	-0.338490159	Lyn	-0.274201146	Fads1	-0.240597737	Rnf20	-0.213142412	Jmy	-0.193156351	Pkrinq1	-0.177832936	Vps37a	-0.162287692	Sgsh1	-0.149033504	Gpi6	-0.13644351
Tmem47	-0.464183811	Tmem9b	-0.337396107	5031426D15Rik	-0.274128699	Rtn3	-0.240178341	Tcp11l2	-0.212816595	Neo1	-0.193043045	Ergic3	-0.177374018	Fam171b	-0.162280719	Fam76a	-0.148998547	Fdft1	-0.136405095
Tnc	-0.463280366	Syt11	-0.337317732	Vimp	-0.273862303	Blcd2	-0.239818842	Sec62	-0.21277183	Emp2	-0.193027934	Ints6	-0.176705768	E330023G01Rik	-0.162184471	Prickle2	-0.148666977	Bcar3	-0.136315718
Mt1	-0.457781151	Plp4a3	-0.335750611	Gpr98	-0.273793855	Kifadb	-0.239796796	Sox2	-0.212540308	Kcnk2	-0.192932502	Bmpr1a	-0.176596533	Crlf1	-0.162120376	Snhg11	-0.148599752	Idh3b	-0.136168351
Slco1c1	-0.455873219	D14Abb1e	-0.334750531	Itpr2	-0.273450816	2310010M20Rik	-0.238182533	Scai	-0.212428134	Man2a2	-0.192548324	Sidt2	-0.175972418	Pnpt1	-0.162083918	2610507I01Rik	-0.148483303	Phactr1	-0.136050662
Pntx6b	-0.454882618	4930480K15Rik	-0.332400254	Igsf11	-0.272769251	Pcsk6	-0.237104288	Cox10	-0.21206619	Lmcl1a	-0.192118463	A630020A06	-0.175831763	Sec24b	-0.161706057	Zfpm2	-0.148145157	Rr45s	-0.136045605
Gabrb1	-0.453287467	Mast1	-0.331618157	Dynll1c	-0.272525383	Lpin1	-0.236310532	Pfdl	-0.211796573	Abcc5	-0.191896602	Agrn	-0.17525745	Cdsn1l	-0.161658326	Rpl6kb1	-0.147946634	Sun4	-0.135974362
Sox9	-0.452373237	Bhlhe41	-0.330387844	Clk1	-0.27242472	Phgdh	-0.23622671	Lrrn1	-0.21724155	Oqt	-0.19184576	Cd38	-0.175157378	Lphn3	-0.161434451	Tspan12	-0.147856894	Fam213b	-0.135643667
Ramp1	-0.44913814	Serpine2	-0.33002277	Tmcc3	-0.272239127	Ncan	-0.236216834	Atm	-0.21130145	Dcaf6	-0.191842395	Nefb	-0.175104813	Scp4	-0.161401968	Cadm1	-0.147789585	Trim12c	-0.135542498
Nfix2	-0.444700077	Kcna2	-0.329656801	Sc4mol	-0.272196934	Sybu	-0.236066798	2310044G17Rik	-0.21130918	Mapk10	-0.19172691	Snx1	-0.175068267	4932419N23Rik	-0.161159812	Rabac1	-0.147331813	Prelp	-0.135346768
Cd63	-0.444176418	Zfp71-rs1	-0.328724872	Mmd2	-0.27189919	Ccdc80	-0.235891454	Eva1a	-0.21170138	1810013L24Rik	-0.191420254	Tmem70	-0.175063411	Ctsa	-0.161127118	Zfp869	-0.147300498	Amd2	-0.135180893
Synj1	-0.444141747	Cyp2j9	-0.328069095	Ctsl	-0.271692818	Gnas	-0.235764665	Tspan15	-0.2106757	Araf	-0.191395987	Nxk6	-0.175012089	Cvxfc	-0.160999941	Ern2os	-0.14685148	Nrp1	-0.135170989
Cd9	-0.440434404	Thbs4	-0.327684294	Ddx17	-0.270464898	Sos1	-0.235710585	Cyb5d2	-0.210503346	Abcc2	-0.19133471	Slc38a1	-0.175005669	Akr1b10	-0.160762589	Ucid1	-0.146804916	Xab2	-0.135123789
Ahcyl1	-0.442468106	Bcl2	-0.327513672	F3	-0.270413316	Rab40b	-0.235654077	Slc1a4	-0.210062306	Zfp799	-0.191097403	Uqp2	-0.174663402	Dpp7	-0.160627769	Mut	-0.146455475	Enpp2	-0.135054408
Fam213a	-0.442055401	S100a6	-0.327507493	Acadl	-0.270243024	Rsph9	-0.234928949	Nadk1	-0.209851492	Vps33a	-0.191051357	Dazap2	-0.174659859	Cacng5	-0.160462917	Syt7	-0.14622141	Emc3	-0.135037838
Gpr37l1	-0.439416678	Srpx	-0.326904425	Sox2ot	-0.270203735	Timp2	-0.234693473	Dpy19l3	-0.209737407	Imr	-0.190789051	2810006K23Rik	-0.174629226	Lrp4	-0.160380511	Ppm1b	-0.146219859	Dcaf7	-0.134975955
Akdar1	-0.434659181	Kcnq1ot1	-0.325928593	Vmn1r58	-0.269913801	Ak1	-0.234360332	Sort1	-0.20938082	Galnt16	-0.190769802	Tnks2	-0.174554419	Napl15	-0.160370068	Ifng2	-0.1455766	Ppp1r12a	-0.134888983
Gltc	-0.432404616	Slc20a2	-0.325914544	Atp8a1	-0.269473106	A630072M18Rik	-0.23429156	Dhrs7	-0.209116157	Trnt1l	-0.19074317	Mbl1	-0.174436358	Prkce	-0.160332692	Pgk1	-0.145338404	Apip	-0.134784227
Calcrl	-0.430943117	Epsti1	-0.322688898	Oat	-0.269394713	Smk	-0.233383077	Ukvld	-0.209069504	Tm2d1	-0.190643505	Wipf3	-0.174322734	Pdf	-0.160232241	Chsy	-0.145305894	Taf15	-0.134741357
Nrcam	-0.430124881	Adcyap1r1	-0.322412783	D830005E20Rik	-0.268695617	Sidt2	-0.232078163	Sypl	-0.209040178	Ctsd	-0.190387243	5730455P16Rik	-0.173955597	Arx	-0.160083136	Crebrf	-0.145100426	Grina	-0.134575949
eCFP transcript	-0.425946595	Slc27a1	-0.321771188	Mxm2	-0.268027279	Ank2	-0.232076188	Nfu1	-0.208900632	Sypl1	-0.19032246	1700128F08Rik	-0.173735617	Cog3	-0.159975824	Pcnp4	-0.145024407	Rab10	-0.134523995
Cmtm5	-0.424467282	Ldhb	-0.319834104	4921508A29Rik	-0.267613665	Pkehbt1	-0.230496393	Clasp2	-0.208640864	Trappc9	-0.190225593	Tmem194b	-0.173247966	Anxa7	-0.159956845	1500015O10Rik	-0.144841042	Jak1	-0.134383535
Slc39a12	-0.424310395	Ulf1	-0.319074717	Rasa2	-0.267275165	Nr2e1	-0.230343291	1110008F13Rik	-0.208461861	Nckap5	-0.188938273	Dnahc7b	-0.173008584	Tkd2	-0.159936402	Zaef1	-0.144835111	Tmem66	-0.134256376
Hspa2	-0.42279469	Gria2	-0.318602887	Gusyfa2	-0.266645001	Ptch1	-0.230270774	Cdkl3	-0.207965084	Tpp2	-0.188839229	Pxdc1	-0.172985982	Lrig1	-0.159815659	Tmem59	-0.144786822	Owsp1a	-0.13421444
Lamc2	-0.415306297	Mgll	-0.318441104	Ddhd1	-0.266340401	Tcf7l2	-0.230201921	Islr	-0.207593502	Rbm28	-0.188662496	Psmb10	-0.172850944	Rdh18l-ps	-0.15981296	Tex2	-0.144754818	Fgd6	-0.134031895
Ertpd2	-0.414991396	Dus3l	-0.318284113	Tppp3	-0.26590084	Stxn2	-0.230173576	Clcn3	-0.206921746	Sash1	-0.188306156	Gstp2	-0.172566009	Mtdh	-0.159772606	Tspan9	-0.144649425	Bcap31	-0.134004495
Aldoa	-0.412488486	Id3	-0.317723153	Sptbn1	-0.265830908	Cd200la	-0.23017017	Fundc1	-0.206860151	Gcc2	-0.188034033	Etl4	-0.172375566	Elf2c3y	-0.159354547	Scrt1	-0.144622673	A330069E16Rik	-0.133754981
Cyp1b1	-0.411051982	Ptxip1	-0.315857252	Gpc5	-0.265506206	Magi3	-0.22963745	Colec12	-0.206709291	Yipf6	-0.187642975	Sec16a	-0.171803517	Xrcc5	-0.159384936	Zfp951	-0.144612048	Suh3	-0.133428367
Omg	-0.410695588	Sfxn1	-0.315162276	Tpp1	-0.264998083	Wdr7	-0.229232864	Nr3c1	-0.20660539	Cbs	-0.18763475	Pnpla7	-0.171465047	Soat1	-0.159055872	Pla3	-0.144572472	Ephb1	-0.133373446
Acot6	-0.4101929	Fam13c	-0.314129735	Atp6v0c-ps2	-0.264721701	Kiflb	-0.229136351	Dhdh	-0.206578104	Klhl	-0.18757754	Scfd	-0.171444891	Arhgef1	-0.159011677	Mlxipl	-0.144563054	Pcnt	-0.133212414
Pla2g16	-0.407841883	4930517M06Rik	-0.313255326	Slc9a3r1	-0.264541606	Slc39a1	-0.229085085	Slc7a4	-0.206465239	Pkrfve	-0.187366092	6720401G13Rik	-0.171418225	Sqstm1	-0.158823882	Rftn2	-0.144146847	Gtf3c1	-0.133129799
Pdcd4	-0.405599027	Kcnj16	-0.313219686	Atp2b1	-0.263232207	Rilact	-0.228976427	Rnf34	-0.206207935	Eya1	-0.187335629	Phactr3	-0.171401054	Mettl15	-0.1587308	Tcoal1	-0.144130462	Pi4ka	-0.133034069
Nxn1	-0.405555059	R3hdm4	-0.312988457	Zfp3612	-0.263199772	Jmjd1c	-0.228896092	Slc25a12	-0.20616022	Tspan15	-0.18732299	Arc1	-0.170938505	Scd1	-0.158700727	Tmem51	-0.144125753	Tmem131	-0.133877131
Gstm5	-0.402993135	Pknb1	-0.31475964	Tcn2	-0.263134447	Mgat1	-0.228894671	Letm2	-0.206133212	Fam86	-0.187303851	Gstk1	-0.170726904	Efha2	-0.158610703	4833422C13Rik	-0.143753148	Efcab14	-0.132860047
Smpd3a	-0.402567058	Hrs1	-0.309794397	Slc35d1	-0.262613491	Cthrc1	-0.228477501	Dio2	-0.206060696	1810037I17Rik	-0.187233427	Atpaf2	-0.170706101	Fktn	-0.158548859	Ncdn	-0.143627713	Cnot2	-0.132569504
Axl	-0.402248883	Hes1	-0.309371712	Fgf2	-0.262477117	Kbtbd3	-0.228246832	Zdbf2	-0.205946717	Lhfp	-0.187062996	Rasip1	-0.170510406	Fam13a	-0.158457752	Pak3l	-0.143539981	Dmtn	-0.132509708
Slc6a1	-0.401153389	Tmem170	-0.309301858	Slc15a2	-0.261871935	Fahd2a	-0.228193805	Msob	-0.205579435	Cimn	-0.186928975	Nptn	-0.170380396	Ssr3	-0.158218771	Tgif	-0.143532829	Arhgap44	-0.1324888
Fxyd1	-0.400106488	2310022B05Rik	-0.308296325	Srxn1	-0.261792673	Nanog	-0.228168779	A830082K12Rik	-0.205273232	Fos	-0.186448454	Uqcrc1	-0.170361165	Poir1b	-0.157820939	Ngn1	-0.143526297	Kdelc2	-0.132387977
Slc4a4	-0.39916309	Cbln2a	-0.307401237	Cdc2	-0.261670422	Zmym6	-0.228106619	Idh1	-0.205035385	Gdi1	-0.186420185	9330159F03Rik	-0.170180347	Tmod1	-0.157761456	Pcmt	-0.143468478	Aimp2	-0.132212427
Agt	-0.399029571	A430107P09Rik	-0.304990276	Tsc22d3	-0.261526155	Med13	-0.22795357	Ppp2r5a	-0.205003284	Pfkm	-0.186358691	Tmem158	-0.170169118	Idi1	-0.157577393	Ythdf3	-0.143282425	Nktrast	-0.132198484
Fut8	-0.397373907	Irf3	-0.303450711	Tacc1	-0.261410306	E130102H24Rik	-0.227846888	Ripply3	-0.204939109	Churc1	-0.186156615	Sh3gl2	-0.170162669	Gpx4	-0.157388265	Eif2s1	-0.143175109	Camta2	-0.132099953
Gm13	-0.396752253	Mcl2	-0.30340234	Gpd1	-0.261164725	Fbxl1	-0.227324381	Tpra1	-0.204795469	Gpx8	-0.185956168	Nfdip2	-0.170156524	Clp4	-0.15746867	Arap2	-0.143161582	Eng5	-0.132049267
2610017I09Rik	-0.396713555	Tmem100	-0.303367112	Dennd1a	-0.261031767	Pon2	-0.227225069	Tfk3	-0.204355533	Tmem59l	-0.185912975	Oaf	-0.169948492	Sgsm1	-0.156866981	Spcs2	-0.143139568	Kcnb1	-0.13190987
Neat1	-0.396652609	Slc24a6	-0.302965948	Wsn2	-0.26077929	Fkbp9	-0.227177635	Mpp3	-0.203921731	B3galt1	-0.185866709	Fam210b	-0.16993665	Dcaf11	-0.156801746	Agpat6	-0.143113454	Zfp617	-0.131815238
Syn2bp	-0.395907591	Slc25a18	-0.302320851	Phyhipl	-0.260765194	Etoh2	-0.226985573	Cwf	-0.203634722	Mettl21c	-0.185799482	Atp6v0c	-0.169880287	Ube3	-0.156750082	Dnlf	-0.143101886	Adcy2	-0.131676186
Kcnj10	-0.395161038	Frzb	-0.301990283	Jam3	-0.260654808	Rab18	-0.225877194	Dmtf1	-0.203254067	Tbk1	-0.185766835	Ankzf1	-0.169748805	Dnpep	-0.156573832	Adck4	-0.143023136	Sfs1l8	-0.131653165
Mfge8	-0.393867387	Cdh4	-0.301957897	Srx14	-0.260575725	Ppp2r2b	-0.225459764	Ficn	-0.202779385	Spire1	-0.185650499	Mbtd1	-0.169482026	Slc13a5	-0.156429787	Lman2	-0.14300781	Fmn4a	-0.131530747
S100a1	-0.393852743	Abca8a	-0.301465674	Mpzl1	-0.260217718	Tmbim6	-0.22570193	Itgb3	-0.202339648	Nxf2	-0.185638587	Rasip1	-0.169445961	Zfp280c	-0.156378916	Zufsp	-0.142954807	Cdh2	-0.13107442
Cpne8	-0.393234431	Rgcc	-0.300969034	Ptprs	-0.260150015	Fam160b2	-0.225079981	Tjp2	-0.202301063	Zfp488	-0.18559856	Lrrc1	-0.169320913	Nav2	-0.15603286	Thrsp	-0.142718272	Gnat3	-0.130920851
Npas3	-0.392242894	Leng8	-0.300545097	Pcdh9	-0.259488216	Clstn2	-0.225042107	Hexb	-0.202272914	Pip1	-0.185507853	Dmd	-0.169220559	Pei2	-0.155903693	Pter	-0.142693217	Rxrer2	-0.130863368
S100a16	-0.39282768	BC033915	-0.300288047	Mkl2	-0.259188689	Lamp1	-0.224667728	Ptpra	-0.202257703	Atraid	-0.185152825	Trim9	-0.169093107	Cyba	-0.155650414	Arfip1	-0.142659963	Zcchc18	-0.130847466
Sowaha	-0.38772285	4933409K07Rik	-0.29950195	Cnp	-0.259014976	Pou3f3	-0.224623475	Canx	-0.20740174	Gabarapl1	-0.185057794	Xrcc4	-0.168846842	Cbfa2i2	-0.155585807	Mpp6	-0.142651527	Eglr2	-0.130759863
Fam20a	-0.386356666	Pde4b	-0.298495065	Pygb	-0.257661254	Bzrp	-0.22447732	Cyr61	-0.201590482	Rnf13	-0.184797041	Ankle2	-0.168814023	Nme3	-0.155527714	Slc22a17	-0.142599756	Axin1	-0.130542944
Acsl3	-0.386192253	Selm	-0.29708797	Manea	-0.257447782	Zscan26	-0.224427517	Ufy	-0.201294362	Itgb3bp1	-0.184794734	Kctd1	-0.168810858	Itsn1	-0.15479473	Sphn	-0.1424796	Vcam1	-0.130368242

Table 3. 2. List of UP TFs and DOWN TFs.

UP TFs	DOWN TFs
Hmgb2	Id4
Hmgb1	Hopx
Sox11	Sox9
Ybx1	Npas3
Nfib	Bhlhe41
Nfix	Zfp71rs1
Insm1	Id3
Hmgb3	Hes1
Ssrp1	Irf3
Smarcc1	Tsc22d3
Zfp367	Rfx4
Smarce1	Nr2e1
Mxd3	Tcf7l2
Zfp62	Pou3f3
Zeb2	Zscan26
Nfia	Dbx2
Sox4	Sox2
Terf1	Nr3c1
Tcf4	Zbtb4
Tulp4	Tfe3
Emx1	Dmtf1
Hmga1-rs1	Bcl6
Zfp637	Fos
Mycn	Ikzf2
Foxg1	Nfx1
Tcf12	Zfp740
Cebpg	Grhl1
Eomes	Klf15
Zmiz1	Zbtb26
Nfyc	4932411N23Rik
Zfp386	Arx
Zkscan1	Etv5
Bcl11a	Mier1
Zfp763	Fezf2
Zfp944	Zfpm2
Zbtb41	Zfp869
Trp53	Lef1
Csde1	Thap7
Zmiz2	Mxi1
Zfp808	Camta2
Zbtb38	Zfp617
Atf4	

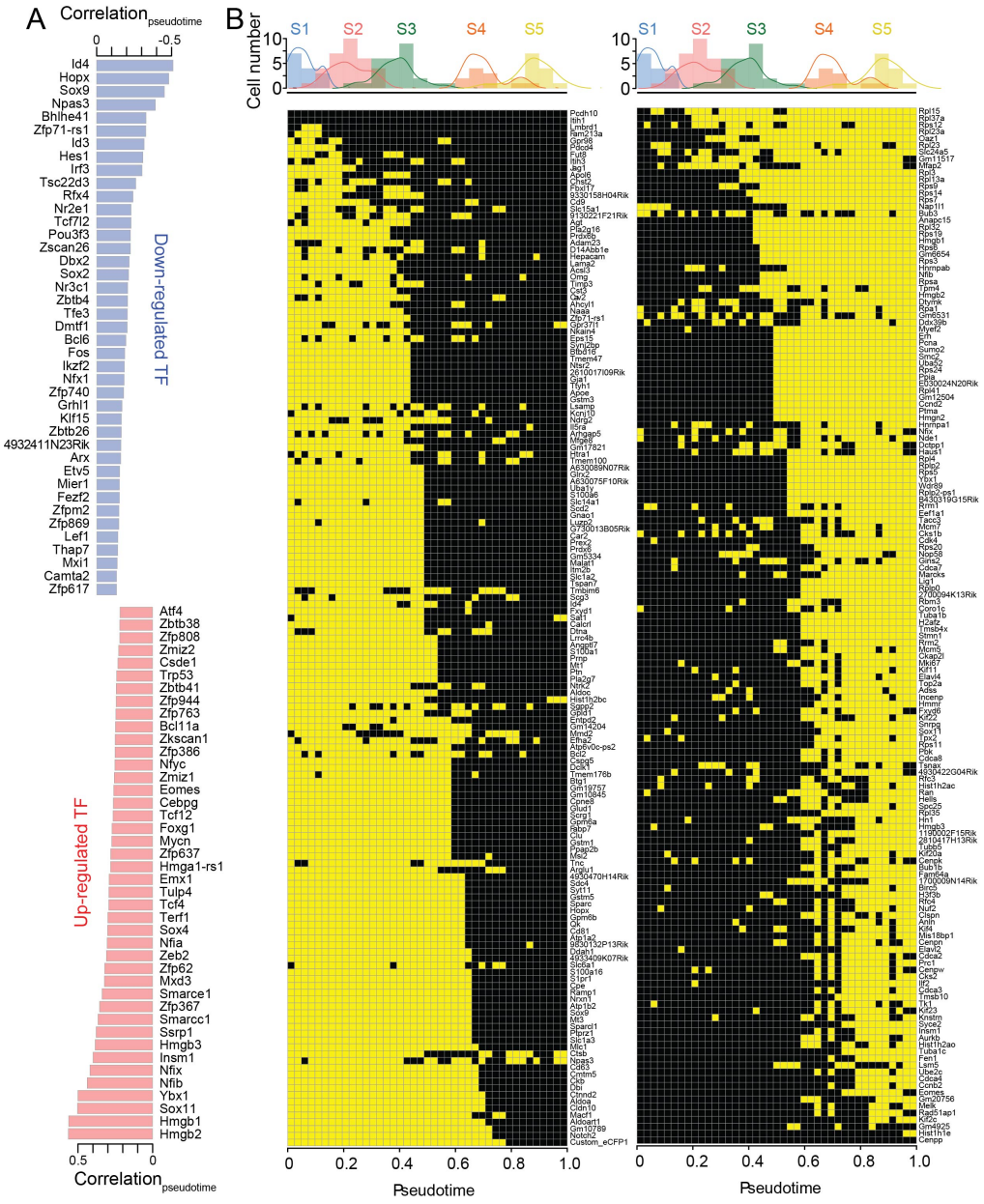

Figure 3. 1. Molecular cascade underlying adult quiescent neural stem cell activation and neurogenesis initiation.

(A) Lists of DOWN and UP TFs and their Spearman correlation coefficient with pseudotime.

(B) ON/HIGH (yellow) or OFF/LOW (black) states of top 150 DOWN (left) and UP (right) genes sorted by the timing of transition points. Shown on the top are histograms of the numbers of individual cells examined along the pseudotime progression. The colors on the histogram follow the color scheme of Figure 2. 3.

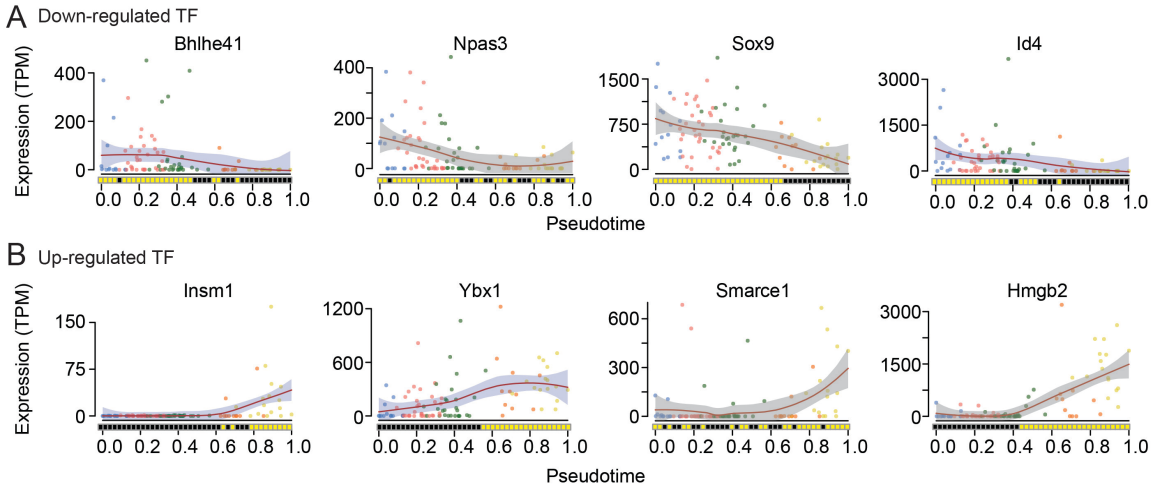

Figure 3. 2. Pseudotime profiles of sample transcription factors.

Shown are pseudotime profiles of representative transcription factors that were down-regulated (A) or up-regulated (B) over developmental pseudotime progression.

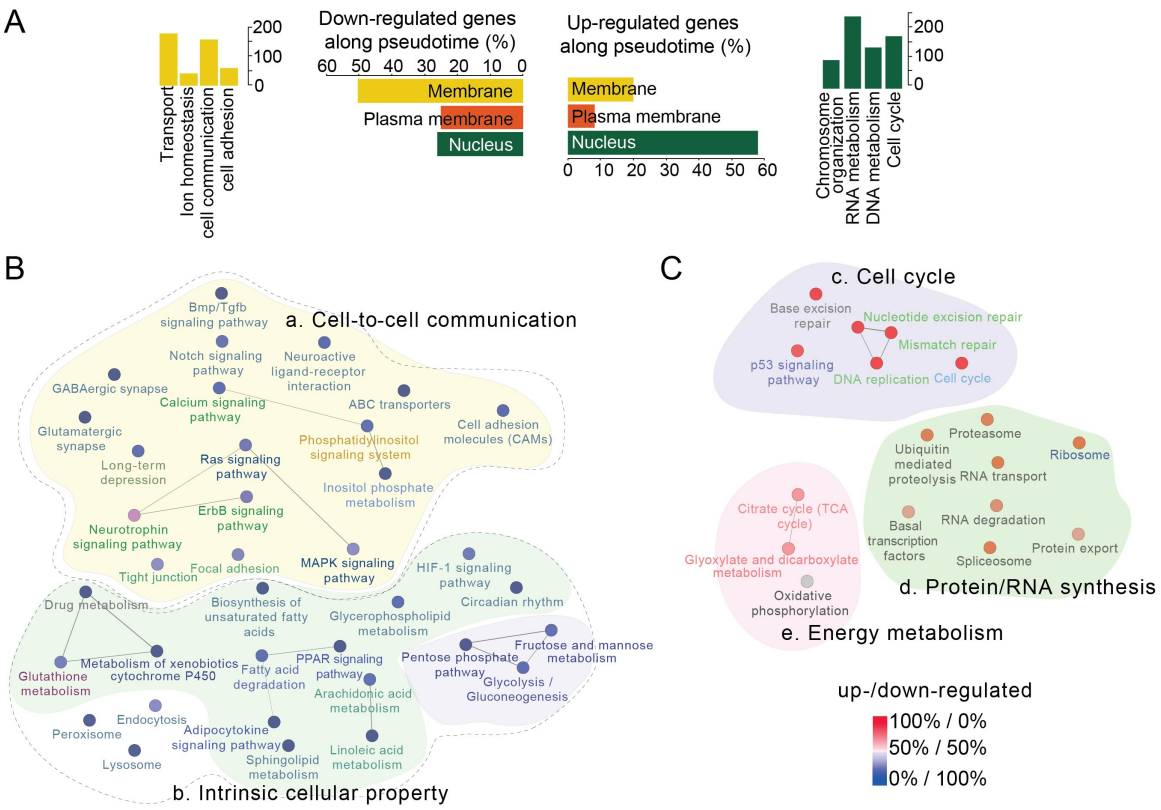

Figure 3. 3. Functional characterization of UP1000 genes and DOWN1000 genes.

(A) Quantification of predicted cellular location of gene products of UP^{1000} genes and $DOWN^{1000}$ genes (two middle panels). Also shown are numbers of genes with indicated functions for membrane-associated DOWN genes (left panel) and those for nucleus-associated UP genes (right panel).

(B) Functional GO analysis for $DOWN^{1000}$ and UP^{1000} genes. Color of each functional entity represents proportion of UP genes (blue) and DOWN genes (red). Connections between each pair of data points represent sharing more than 5 genes between the pair. Functionally similar entities are grouped with same background colors. Broken lines represent two categories of DOWN genes: genes encoding proteins involved in the intra- or extra-cellular communication and genes encoding proteins defining intrinsic stem cell properties. The P values are from hypergeometric tests, and corrected by Holm–Bonferroni method.

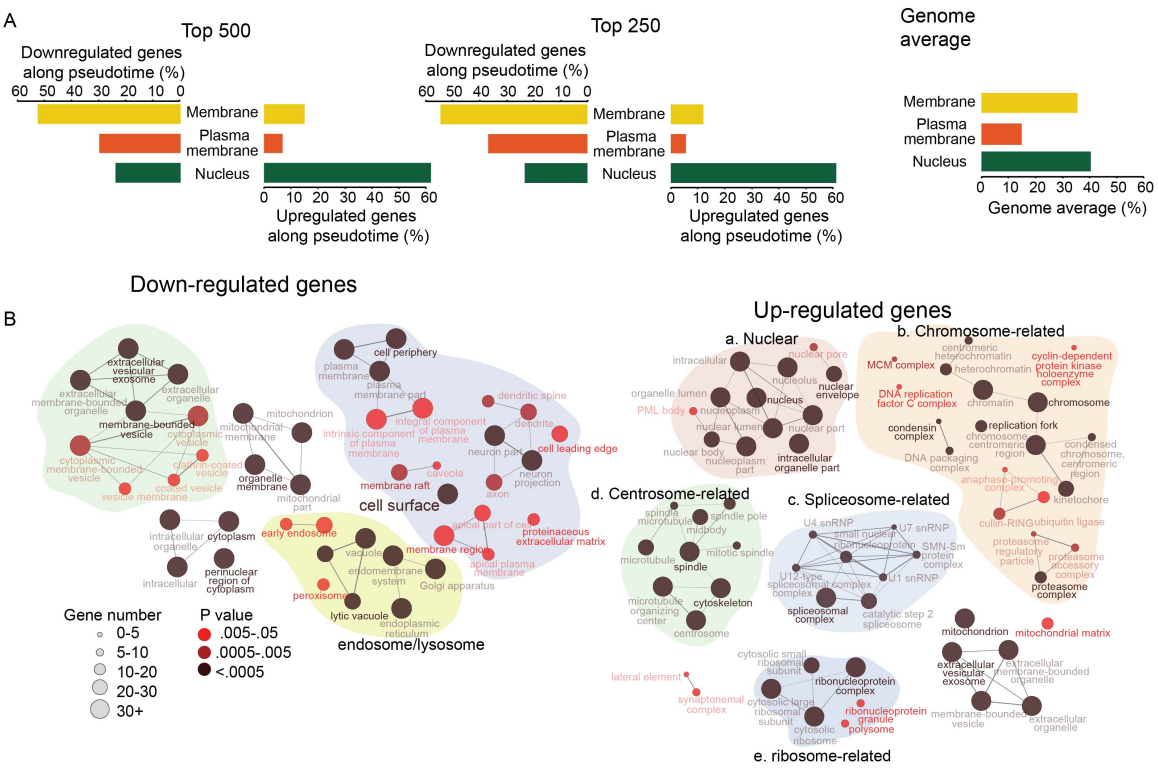

Figure 3. 4. Cellular compartment analysis for UP and DOWN genes.
(A) Cellular compartment analysis for UP and DOWN genes with different thresholds. Shown are summaries of enrichment patterns of UP and DOWN genes with cutoff for top 500 genes (left panel) or top 250 genes (middle panel), and average for all genes (right panel).

(B-C) Detailed predicted cellular locations for gene products of DOWN^{1000} genes (B) and of UP^{1000} genes (C). The size of each data point represents the number of genes within each predicted location, and color of each data point represents P value of enrichment to each predicted location. P values are from hypergeometric test and corrected by Holm–Bonferroni method (Bonferroni step-down correction). Connections between pairs of data points represent sharing more than 50% of genes between the pair. Similar entities are grouped with the same background colors.

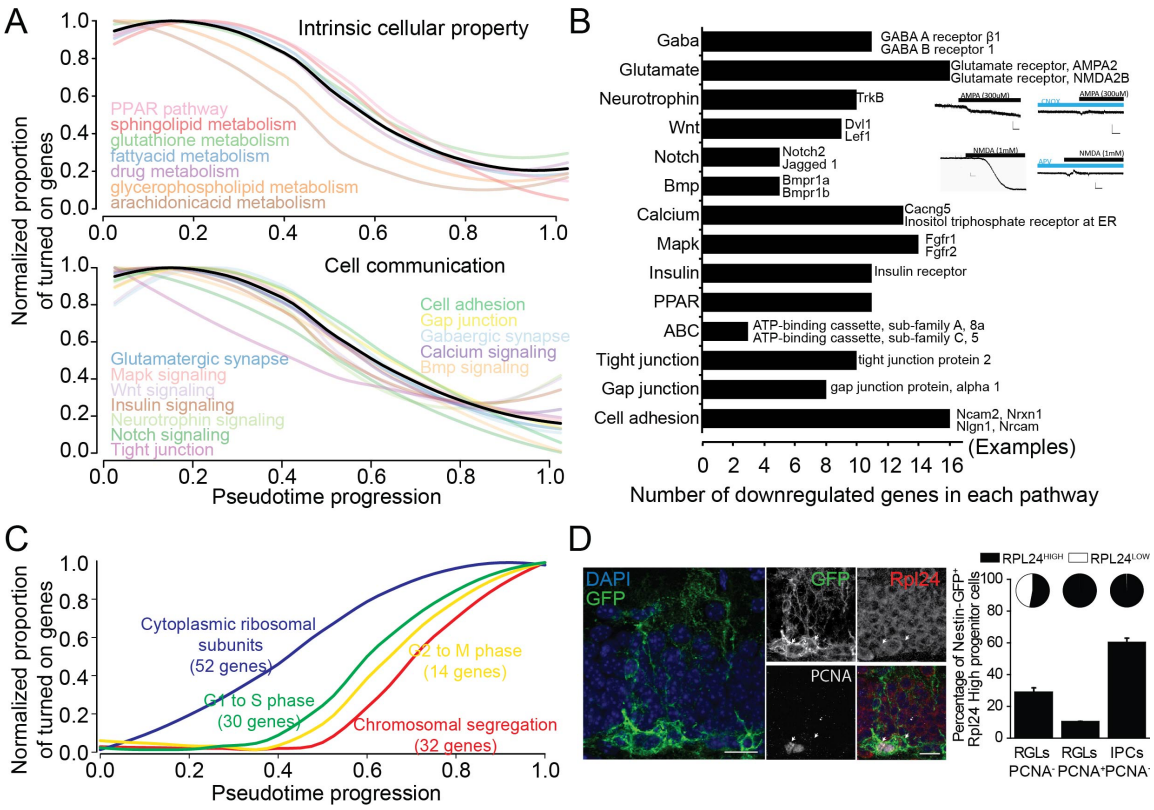

Figure 3. 5. Sequential molecular dynamics during adult neural stem cell activation and neurogenesis.

(A) Gradual down-regulation of averaged binary states of each functional entity defining intrinsic stem cell properties (top) and those defining intra- or extracellular communication (bottom). On/off binary states of expressed genes within each functional entity were averaged and normalized to show the timing of the transition.

(B) Number of genes in different DOWN gene ontology groups and representative example genes. Expression patterns of representative genes over developmental pseudotime are shown in Figure 3. 6A. Electrophysiological recording of Nestin-GFP^{cyto+} NSCs in acute hippocampal slices showed responses to both AMPA and NMDA (bottom).

(C) Gradual up-regulation of averaged binary states of each functional entity defining cell cycle checkpoints and cytoplasmic ribosomal subunits. On/off binary states of up-regulated genes within each functional entity are averaged and normalized to exemplify the timing of the transition. Cell cycle checkpoint genes were up-regulated in the sequence of the cell cycle progression. The up-regulation of cytoplasmic ribosomal subunits preceded up-regulation of the earliest cell cycle checkpoint gene.

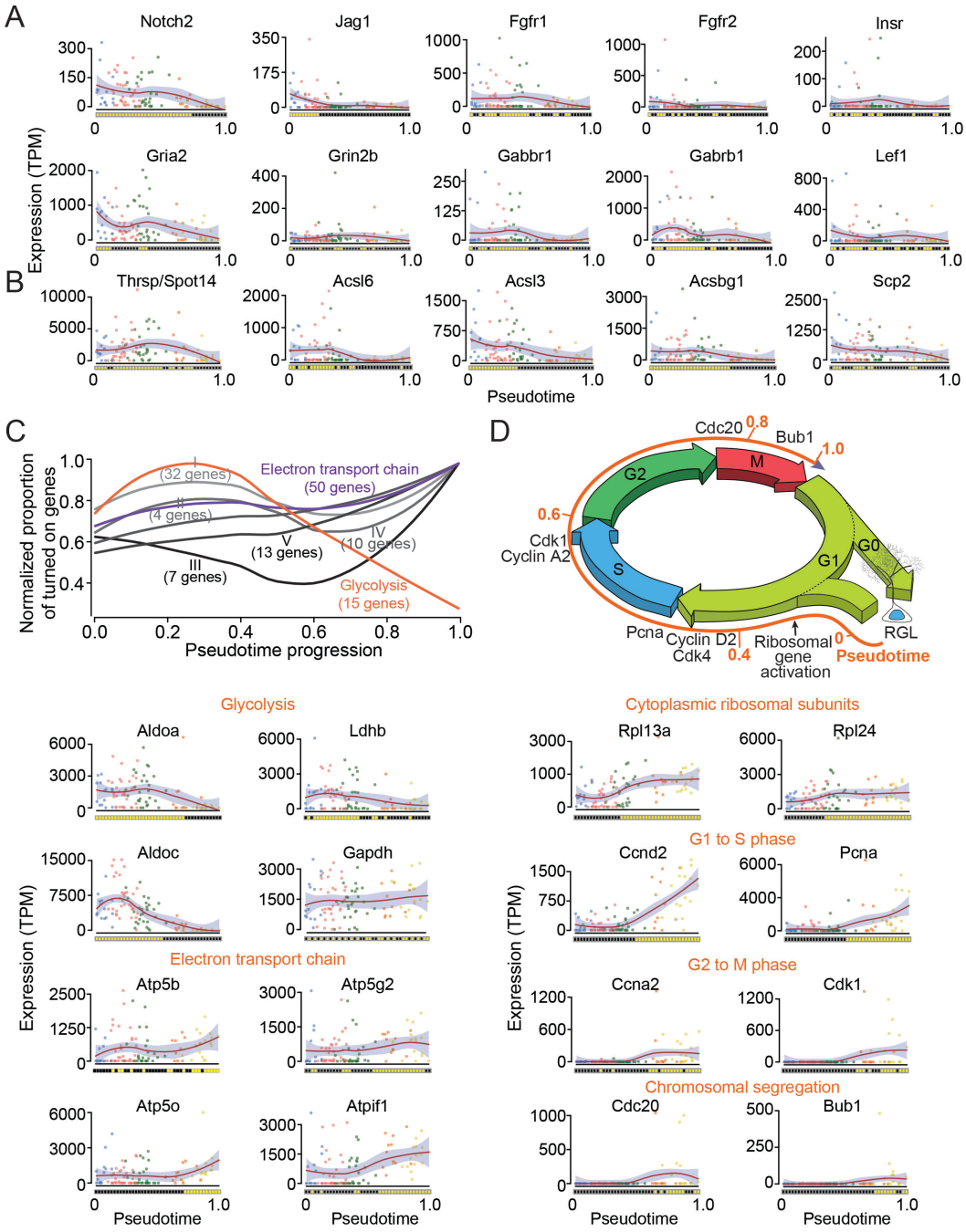

Figure 3. 6. Representative pseudotime profiles of key genes related to niche signaling, cell cycle progression, and energy metabolism.

(A-B) Pseudotime profiles of representative genes related to niche signaling (A) and genes related to fatty acid metabolism (B), both of which exhibited down-regulation upon qNSC activation. Similarly plotted as in Figure 2. 3D.

(C) Reciprocal changes of expression of mitochondrial electron transport genes and glycolysis genes. Roman numerals represent electron transport complex. Complex V exhibited a clear increase throughout early neurogenic pseudotime. Shown at the bottom are pseudotime profiles of some genes related to glycolysis (gradual down-regulation) and complex V genes in mitochondrial electron transport chain (gradual up-regulation).

(D) Developmental pseudotime recapitulates cell cycle progression during qNSC activation and initiation of neurogenesis. Shown on the top is a schematic illustration of cell cycle phases correlated with pseudotime. Cell cycle checkpoint genes were up-regulated following the sequence of biological cell cycle progression. Shown at the bottom are pseudotime profiles of genes related to cell cycle checkpoints and cytoplasmic ribosomal subunits. Note that ribosomal genes were up-regulated earlier than any of the cell cycle related genes.

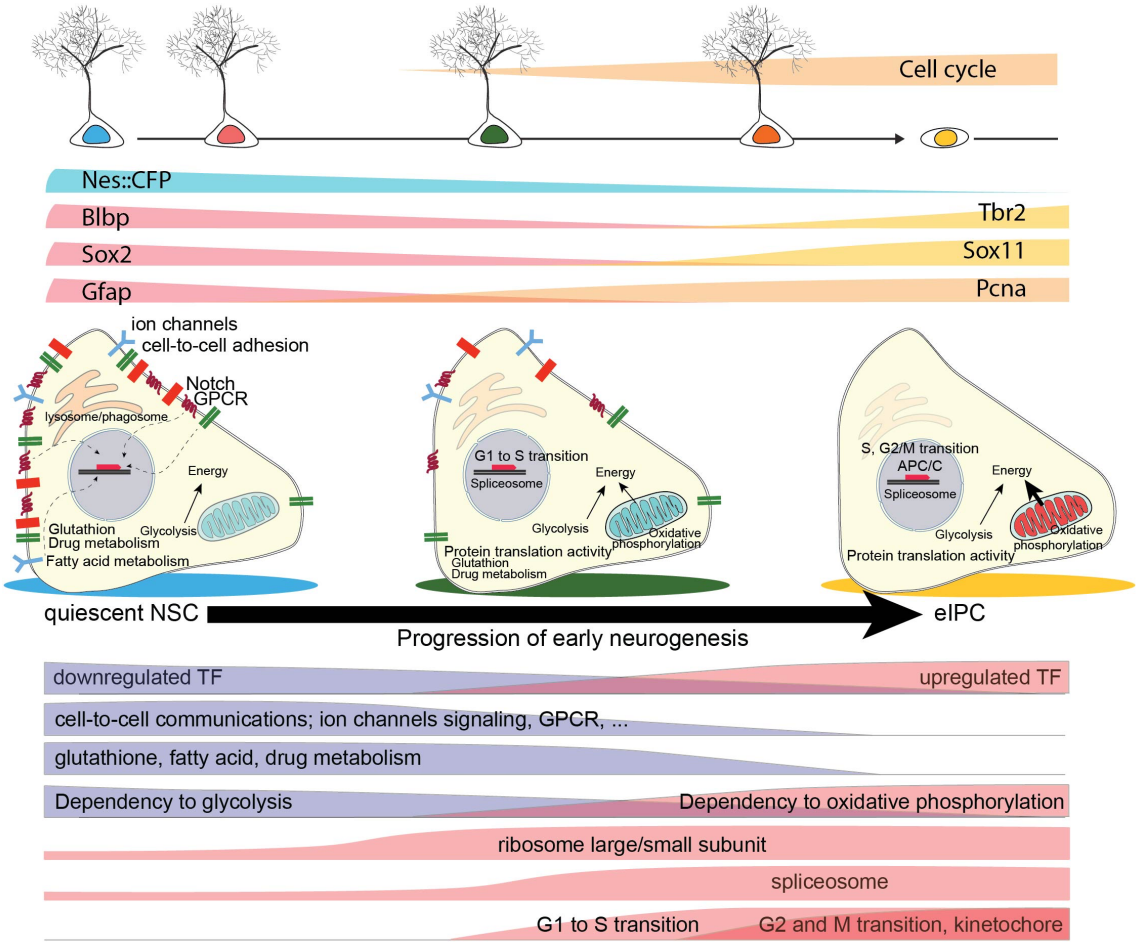

Figure 3. 7. Schematic summary of molecular signatures of quiescent adult neural stem cells and molecular cascades underlying their activation and neurogenesis.

Shown on top is an illustration of known marker expression and cell cycle activation during adult hippocampal neurogenesis. Shown in the middle is an illustration of molecular signatures of adult qNSCs and their immediate progeny. Shown at the bottom are functional categories of genes that show a clear shift during adult qNSC activation and generation of eIPCs. qNSCs exhibit intra- and inter-cellular signaling to actively sense the local niche, rely mostly on glycolysis for energy, and have highly active fatty acid, glutathione, and drug metabolism. Upon activation, NSCs increase translational capacity, followed by cell cycle entry with G1 to S transition. Oxidative phosphorylation starts to be active and stem cell specific properties are down-regulated. eIPCs maintain active cell cycle genes, ribosomal activity and fully active oxidative phosphorylation for energy generation. The color scheme on the top and the middle illustration follows the colors from Figure 2. 3B.

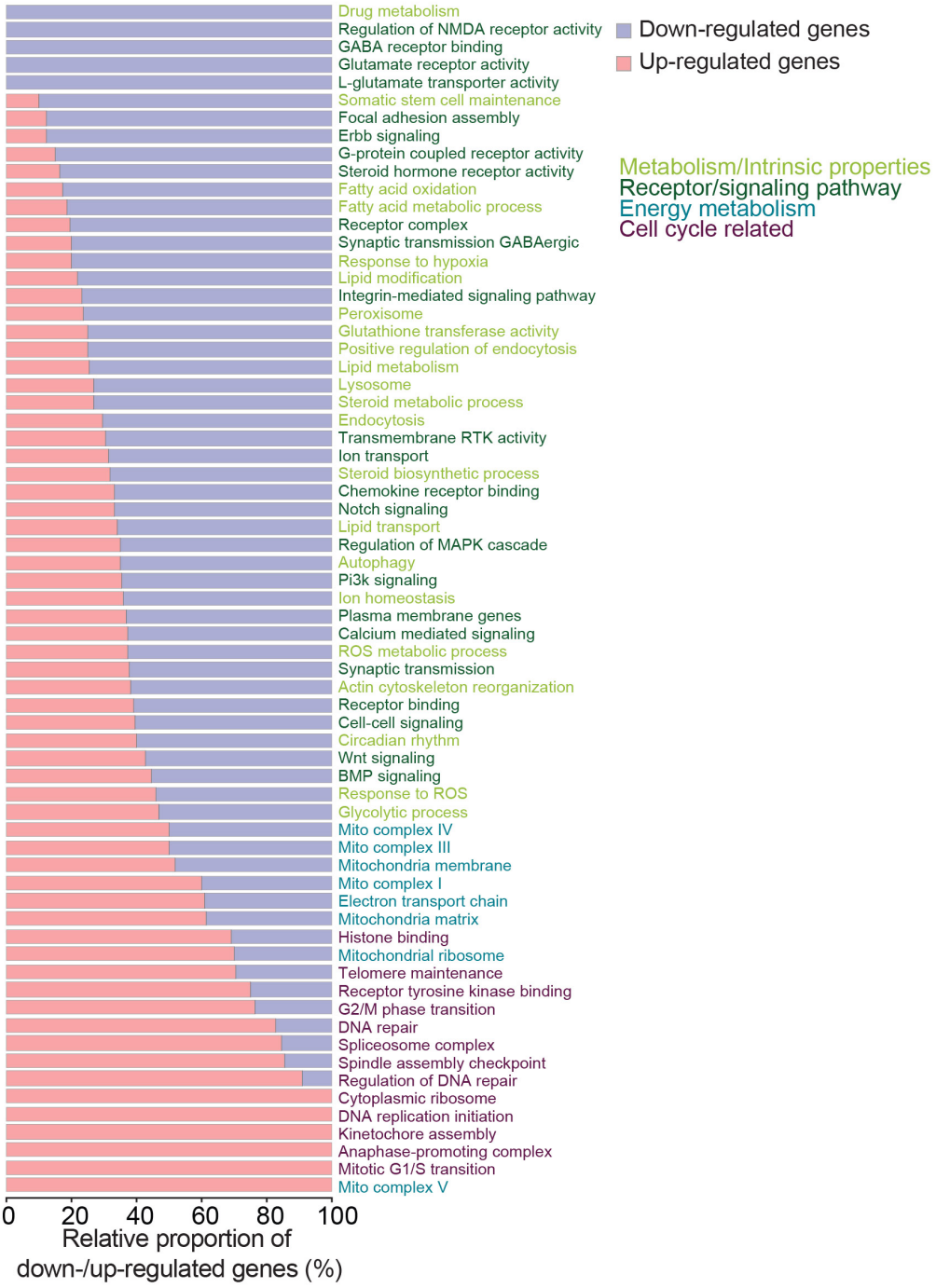

Figure 3. 8. Independent validation for GO entity enrichment test for up-regulated and down-regulated genes during quiescent stem cell activation and neurogenesis.

Shown is a summary of the proportion of up-regulated and down-regulated genes within each key functional GO entity. GO entities with a disproportionally higher proportion of up-regulated genes represent functional pathways that are activated during exit of quiescence and early stages of neurogenesis, whereas GO entities with a disproportionally higher proportion of down-regulated genes represent functional pathways that are qNSC-specific pathways.

Chapter 4. Discussion[4]

[4] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

Understanding adult NSC behavior and neurogenesis requires quantification of molecular states along a continuous developmental process. In the current study we generated three major resources. First, we provide a comprehensive dataset of single-cell transcriptomes of qNSCs and their immediate progeny in the adult mouse hippocampus in vivo. Second, we provide Waterfall, an unsupervised bioinformatic suite for in silico reconstruction of molecular trajectories based on snapshots of single-cell transcriptomes and statistical gene expression analysis over continuous developmental processes. Third, we provide a holistic picture of adult qNSC molecular signatures and dynamic molecular cascades underlying initial phases of adult neurogenesis at unprecedented temporal resolution (Figure 3. 7). Our study provides an example of how to resolve cellular heterogeneity and reveal developmental dynamics for systematic molecular characterization of stem cells and their differentiation in vivo. Our approach can be adapted for various single-cell omics analyses (transcriptomics, proteomics, epigenomics, lipidomics, and metabolomics) of many continuous biological processes, such as development, physiological and pharmacological stimulation, and disease progression (See examples in Chapter 5. Single cell analysis with Waterfall).

A Resource of in vivo Single-cell Transcriptomes of qNSCs and their Immediate Progeny

Our study provides a single-cell RNA-seq dataset and the first comprehensive view of transcriptome dynamics underlying adult qNSC behavior in vivo. Currently, there is no published dataset for transcriptome dynamics during stem cell development in any somatic system in vivo. We performed multiple levels of validation of our dataset and approach, including comparison with an in situ database, confirmation with known and unknown marker expression during adult neurogenesis in vivo, and functional validation via clonal lineage tracing and electrophysiology (Figures 3 and 6B; Table 2. 2).

Our resource of holistic molecular profiles during early neurogenesishas three unique features that increase its versatility. First, our whole-transcriptome information includes unannotated transcripts, isoforms, and retrotransposon-derived transcripts, as opposed to multiplexed qPCR or microarray-based studies which can only provide limited annotated transcripts (Hoppe, Coutu et al. 2014). Second, we animated static single-cell transcriptomes over

the in vivo neurogenesis trajectory, allowing queries of molecular dynamics for each gene over the continuous biological process and generation of novel functional hypotheses during specific phases of adult qNSC maintenance, activation and neurogenesis initiation. For example, expression of nuclear glucocorticoid receptor Nr3c1 in adult qNSCs but not in eIPCs (Figure 3. 1A) suggests a cellular target of glucocorticoids during adult hippocampal neurogenesis and a means to manipulate qNSCs in vivo. Our resource also reveals potential prospective markers of adult neurogenesis, such as using *Hopx-CreER*^{T2} to target adult qNSCs in vivo (Figures 3E-G). Third, each transcriptome in our dataset represents a biological state within a single cell. This modular construction allows for flexible reorganization of the dataset to probe different questions as our understanding of the biology evolves. For example, instead of generating reporter lines for individual genes or using different surface markers for physical sorting of specific cell populations, investigators can perform unlimited in silico cell sorting with any individual gene, or multiple genes in combination, to obtain a selected cell population to probe their gene expression characteristics at the genome-wide level using our single-cell datasets.

Developmental Dynamics of Adult Neurogenesis at the System Level

Recent genome-wide studies have begun to provide a system-level understanding of in vivo adult NSC biology using marker-defined NPC populations (Kriegstein and Alvarez-Buylla 2009, Bracko, Singer et al. 2012, Codega, Silva-Vargas et al. 2014). Yet previous studies have only acquired snapshots of transciptomes, which limits investigation of developmental dynamics among different cellular states. We co-opted the imperfection of the *Nestin-CFP*^{nuc} genetic labeling system to collect individual qNSCs and their immediate neuronal progeny concurrently. Aligning cells along the developmental trajectory yielded, for the first time, a molecular continuum with sequential progression of the individual transcriptome from qNSC to aNSC and then eIPC. Importantly, this novel approach does not divide developmental processes into discrete stages that are defined a priori by capturing populations sharing specific markers.

Our resources provide unparalleled temporal resolution to identify new mechanisms underlying adult NSC biology. For example, we showed that Acyl-CoA synthetases (Acsl3, Acsl6 and Acsbg1), the enzymes for the first step of fatty acid β-oxidation, were highly expressed only in qNSCs (Figure 3. 6B), suggesting a novel role for active fatty acid β-oxidation in qNSCs and

thereby extending previous findings on the role of fatty acid metabolism in adult neurogenesis (Knobloch, Braun et al. 2013)(Table 2. 2). We also found that ribosomal subunits were the first genes up-regulated upon adult qNSC activation and early differentiation (Figure 3. 5C and Figure 3. 6D), suggesting a possible demarcation of G0 to G1 transition and providing the timing for switches in protein synthesis, the regulation of which is important for somatic stem cell function (Signer, Magee et al. 2014). The holistic picture we obtained unifies disparate information and illuminates novel biological themes in stem cell biology. For example, our analysis suggests that qNSCs actively respond to local environmental cues through various signaling pathways, but gradually and globally shut off signaling capacity upon activation. These observations support the concept of a niche wherein mammalian somatic stem cells are tightly controlled by a regulatory microenvironment (Schofield 1978), and predict that eIPCs are less responsive to environmental input (Berg, Yoon et al. 2015). This novel biological insight may be applicable to many somatic stem systems defined by stochastic behavior (Simons and Clevers 2011).

Waterfall Analysis of Single-cell Transcriptomes within a Continuum

Waterfall has three key differences from previously methodologies. First, it requires very little prior information to generate a highly accurate temporal trajectory at single-cell resolution. Previous methods have been able to reconstruct accurate trajectories by relying on a robust set of known markers to establish cell order and validate cell alignment at numerous points along the timeline. For many biological systems and processes, we have much less information. Second, in contrast to Monocle (Trapnell, Cacchiarelli et al. 2014) or Wanderlust (Bendall, Davis et al. 2014), Waterfall uses k-means clustering to build a trajectory and assign an individual cell a pseudotime based on each cell's proximity to the cluster-derived trajectory, rather than constructing a trajectory by directly connecting each cell to the next. Third, in order to analyze stochastic gene expressions, we adopted HMM to predict consecutive binary states in gene expression activity over pseudotime. HMM permits the interpretation of highly variable data without logarithmic transformation, normalization, or the input of any arbitrary parameters, such as threshold for gene expression noise or Markovian parameters (transition probability, initial probability and emission probability). Our validation for known and unknown genes in adult neurogenesis indicated that HMM correctly predicted temporal dynamics of in vivo biology.

There is no conceptual restriction of our approach to transcriptome studies of adult neurogenesis. Indeed, in Chapter 5. Single cell analysis with Waterfall, we provide examples of how Waterfall could be broadly applicable for single-cell RNA seq datasets such as in vitro myogenesis, in vivo embryonic lung development, single-cell mass-cytometry dataset from in vivo B cell development, and synthetic datasets. We expect that Waterfall algorithms can be adopted for diverse single-cell multi-dimensional datasets, including single-cell transcriptomes, epigenomes, proteomes, and metabolomes, of various continuous biological processes.

Chapter 5. Single cell analysis with Waterfall[5]

[5] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

I. Development of Waterfall

Investigating the molecular signatures of adult hippocampal neurogenesis at the single cell level presents considerable challenges – challenges both endemic to the field of single cell analysis and unique to the adult neurogenesis system. In this chapter, we briefly highlight the most critical of these challenges and our strategy to resolve them. We then provide a detailed summary of our development and application of Waterfall to our adult neurogenesis dataset, followed by a comparison of the results with those obtained by Monocle for the same dataset. We then evaluate how well our bioinformatic strategy handles data from different biological systems, each with its own characteristics and constraints. The datasets we analyzed are shown in Table 5. 1. List of datasets to which Waterfall was applied. .

What is Waterfall?

Waterfall is a bioinformatics pipeline to process multi-dimensional single cell dataset derived from continuous biological processes. Waterfall comprises three steps (Figure 5. 1):

Pre-processing

a. Removing outliers
b. Determining and orienting a single route of interest and constitutive single cell clusters

Reconstruction

a. Reconstruct a trajectory using minimum spanning tree (MST)
b. Determining the relative chronological location (pseudotime) of each single cell

Gene expression analyses

a. List of positively or negatively correlated genes with pseudotime
b. Determining underlying binary gene expression states using hidden Markov model

Challenges of analyzing single cells in continuous biological processes

Using single cell molecular profiles, we can deconstruct heterogeneity at the molecular level with greater precision than is possible from analyzing bulk populations. The goal and methodology of single cell analysis is tightly bound to the type of heterogeneity that to be resolved.

Resolving static or cell-type heterogeneity is analogous to directly comparing group data at the population level, and can be performed with conventional analytical tools. The goal is to identify distinct categories of cells based on the molecular profiles and classify cells accordingly. Recently developed cell-type prediction software has great potential to facilitate these studies (Cahan, Li et al. 2014). In contrast, there are no tools for solving dynamic or temporal heterogeneity that can be adopted from population level studies, because it was not until the advent of single cell technology that we had access to molecular profiles at high enough resolution to track gradual transitions along continuous biological processes. Single cell analyses of dynamic processes face two critical challenges: (1) lack of information and (2) stochasticity of the single cell transcriptome (Figure 5. 2).

Lack of information

In order to deconstruct heterogeneity within a continuous process like development, we need to find genes that change during the process to provide points of reference and alignment (Buettner, Natarajan et al. 2015).

Bendall et al. used known set of markers involved in the biological process, thereby successfully de-convoluting the in vivo B cell developmental process from single cell mass cytometry data (Bendall, Davis et al. 2014) (Figure 5. 2A). All the single cell mass cytometry and single cell qPCR studies use this approach (Bendall, Simonds et al. 2011, Qiu, Simonds et al. 2011, Bodenmiller, Zunder et al. 2012, Guo, Luc et al. 2013, Bendall, Davis et al. 2014, Saadatpour, Guo et al. 2014). Applying this methodology, however, requires a substantial understanding of the biological system, where we know most, if not all, of the important markers to define cell types or developmental stages. For most biological system other than the hematopoietic system, we do not yet have enough information to identify key genes or stratifiers.

Trapnell et al. nicely solved this challenge by discovering key genes through population comparison (Figure 5. 2B) (Trapnell, Cacchiarelli et al. 2014). Trapnell et al. astutely and successfully reconstructed the trajectory of in vitro skeletal muscle differentiation at single cell resolution using differentially expressed genes from population comparison between before and after switching media to initiate differentiation (Figure 5. 2B). To perform this kind of population comparison, however, we need to know the temporal boundaries that define subgroups before analyzing at the single cell level. Such information is a clear advantage of in vitro systems and rarely available for in vivo systems, considering the heterogeneity of cell types and asynchronous cellular states (Figure 5. 2C).

Stochasticity of single cell transcriptome

Single cell gene expression is highly stochastic, whereas the population level transcriptome is more stable due to averaging thousands to millions of single cell transcriptomes (Ramskold, Luo et al. 2012, Shalek, Satija et al. 2013). One of the most widely used models to explain single cell level gene expression is a bistable transcription model, wherein underlying gene expression activity oscillates between an active state and a basal state (Novick and Weiner 1957, Levsky, Shenoy et al. 2002, Ozbudak, Thattai et al. 2004, Lim and van Oudenaarden 2007, Arnoldini, Vizcarra et al. 2014). Assuming the mRNA transcription is a major determinant for mRNA abundance (Schwanhausser, Busse et al. 2011), we could approximate single cell transcriptome stochasticity as the random variability from two disparate underlying gene expression states. When analyzing single cell datasets from continuous biological processes, we need to furthermore consider the transition events between the underlying states during the process. Thus, we need a novel approach that can better reflect the single-cell biology. An ideal approach should explain:

1. Transition events between the two underlying states throughout the continuous process
2. Different points of the transition events of individual genes over the continuous process
3. Different gene expression distributions resulting from the two underlying states
4. Noisy or missing data points

We proposed to use a hidden Markov model (HMM) when interpreting single cell data from continuous processes (Figure 5. 3). Major assumptions of using HMM for single cell data from continuous trajectory are:

1. There are two underlying states emitting different single cell gene expression distributions.
2. The bi-stable state model is the major sources of single cell gene expression stochasticity.
3. The pseudotime bin (Figure 2. 6) is small enough to contain cells with identical “hidden” states for most genes.
4. Gene expression from each state follows a normal distribution.

We believe the first two assumptions are reasonable, as shown by multiple single cell studies (Novick and Weiner 1957, Levsky, Shenoy et al. 2002, Ozbudak, Thattai et al. 2004, Lim and van Oudenaarden 2007, Arnoldini, Vizcarra et al. 2014). The third assumption depends on the size of bins, thus can be iteratively modified through comparison between the prediction and the known biology. The fourth assumption is also reasonable, since we are not forcing the gene expression distribution to a Poisson distribution. In addition, HMM is relatively immune to noisy or missing data points.

Admittedly, HMM might be one of many potential models to explain stochastic single cell RNA-seq data over continuous trajectories. However, there is no attempt yet to try to systematically interpret stochasticity in a single cell RNA-seq dataset. We introduced this concept for the first time and it fits well with the known biology as could be validated for specific genes.

II. Analyzing single cell data using Waterfall

We originally developed the Waterfall pipeline to analyze the single cell RNA-sequencing dataset from in vivo adult hippocampal neurogenesis. In this chapter, we present detailed analytical procedures of how we applied Waterfall for the adult neurogenesis system. Then, we tested how Waterfall performed in other systems (e,g. in vitro myogenesis, embryonic lung development, in vivo B cell development) and platforms (e,g. single cell mass cytometry) (Bendall, Davis et al. 2014, Trapnell, Cacchiarelli et al. 2014, Treutlein, Brownfield et al. 2014). In addition, we compared the performance of Waterfall and Monocle, which is a pioneering and currently the only available software for single cell RNA-seq from continuous processes (Trapnell, Cacchiarelli et al. 2014). Finally, we also applied Waterfall or two synthetic datasets to show its optimal behavior under ideal conditions.

1. Adult neurogenesis dataset (In vivo single cell RNA-seq)

Nestin-CFP^{nuc} is highly expressed in the neural stem cell (NSC) population and is carried over to their immediate progeny (early immediate progenitor cells, eIPCs). Using this imperfect labeling system to our advantage, we were able to capture the continuous cellular states between quiescent NSCs (qNSCs) and eIPCs.

Mapping and calculating normalized expression levels

The CFP^{nuc+} cells expressed higher levels of CFP transcript compared to CFP^{nuc-} cells (Wilcoxon signed-rank test p value = 8.018×10^{-16}) (Figure 5. 4A). The 3' bias of the amplification was comparable to other single cell RNA-seq studies (Figure 5. 4B). Of note, DNase treatment significantly improved the data quality (Figure 5. 4B). We used RSEM to calculate TPM values (Li and Dewey 2011).

Waterfall: Preprocessing

Nestin-CFP^{nuc} mouse system also labels a small percentage of non-NPCs in the adult dentate gyrus (Figure 2. 2A). We excluded cells that either exhibited markedly different transcriptomic profiles from majority of the CFP^{nuc+} population or were identified as non-NPCs, such as oligodendrocyte progenitor cells or pericytes (Figure 5. 5).

We then performed unsupervised clustering and principal component analysis (PCA; Figure 5. 6). S1 to S5 formed a linear trajectory, whereas SA formed a tangential branch (Figure 5. 6B) as also indicated by the unbiased minimum spanning tree (MST; Figure 5. 7A). Gene expression profiles at Figure 5. 6C indicated that the S1-S2 population represents the quiescent neural stem cell (qNSC) population, whereas the S5 branch represents the early intermediate progenitor cell (eIPC) population, which gives rise to newborn neurons in adult dentate gyrus.

Waterfall: Additional trajectories

Although S1 to S5 path represents the most probable path of adult neurogenesis, there are other possible routes to consider (Figure 5. 7). First, we have three hypotheses for the identity of SA (Figure 5. 7C, left and middle): (1) astrocytic differentiation from NSC depletion (Encinas, Michurina et al. 2011)(Figure 5. 7C, left); (2) activated NSCs returning to quiescence after asymmetric or symmetric division (Bonaguidi et al., 2011); (3) a distinct adult NSC population (Figure 5. 7C, middle). SA expressed adult qNSC genes, such as Gfap at similar or even higher level than S1 and S2. They also expressed significantly higher levels of oxidative phosphorylation components (Cyb5, Atp5o; unadjusted p-values 0.0275 and 0.00211, respectively) and mitochondrial ribosomal subunits (Mrps11, Mrps21; unadjusted p-values 0.0156 and 0.0234, respectively), suggesting that SA was at a distinct metabolic state.

Second, while building up a pseudotime sequence, we initially noticed that a few cells located at the beginning of the trajectory expressed lower levels of adult NSC markers (Figure 5. 7C, right). Thus, we excluded the five cells, which were located at the very end of S1 for pseudotime analyses. Our working hypothesis is that they might be exhaustive gliogenic population (Bonaguidi, Wheeler et al. 2011, Encinas, Michurina et al. 2011), since these five cells expressed high level of glial markers. We do not have resolution to draw a line between qNSC populations from potential glial differentiating population. Although these possibilities propose exciting possible directions for more data mining from our single-cell RNA-seq dataset, we decided to focus on the neurogenic trajectory.

Waterfall: Reconstructing trajectory and pseudotime

We used k-means clustering and minimum spanning tree to generate a trajectory (Figure 5. 8A). The number of clusters was determined iteratively using the elbow method (Thorndike 1953). We assumed that the tangential components of single cells to the trajectory were derived from the stochastic gene expression or technical variability. We assigned pseudotime to each single cell based on its relative location over the pseudotime trajectory (Figure 5. 8B).

Waterfall: Gene expression analyses

Pseudotime allows profiling of gene expression dynamics throughout the reconstructed trajectory (Figure 5. 9). As expected, the single cell transcriptome was highly stochastic (Figure 5. 9). The regression plot was not quantitative enough because the shape and deflection point of the regression plot varies greatly based on arbitrary parameters such as degree of smoothing. Moreover, the regression plot does not reflect the stochastic and bimodal nature of single cell expression. HMM is more quantitative and does not require setting any arbitrary parameters (Figure 5. 9, yellow/black block graphs).

Applying Monocle to the adult neurogenesis dataset

We applied Monocle to our adult neurogenesis dataset to compare its performance with Waterfall (Trapnell, Cacchiarelli et al. 2014). In order to perform a fair comparison, we used TPM both for Monocle and Waterfall, because TPM was shown to be more stable in analyzing RNA-seq data (Wagner, Kin et al. 2012). We then removed oligodendrocytes, pericytes or other unidentified outliers based on their marker expression before applying Monocle (Figure 5. 5).

In vivo derived single cell data do not entail spatiotemporal information necessary to select for key developmental genes to stratify the single cells. Since Trapnell et al. recommends selecting highly expressed genes, we chose a trajectory with top 1000 genes (Figure 5. 10).

We compared the gradual transition of single cell transcriptome along the trajectories determined by Waterfall and Monocle, which was evaluated by the sum of the distance between each pair of neighboring cells from the beginning to the end of the pseudotime sequence (Figure 5. 11A). The more gradual the trajectory, the shorter the total distance would be. As expected, Waterfall and Monocle resulted in significantly shorter distances compared to a randomly permuted cell order (Figure 5. 11B). Further, the total path distance of the pseudotime by Waterfall was significantly smaller than the pseudotime by Monocle. The cumulative probability in Figure 5. 11B demonstrates that it requires ~10^{18} trials of random permutation of single cell orderings to achieve the precision of Waterfall and ~10^{10} trials of random permutation of the single cell orderings to achieve the precision of Monocle.

Monocle showed expected profiles of qNSC-enriched genes (Apoe, Fabp7, and Gfap) and eIPC-enriched genes (Tbr2/Eomes and Sox11) (Figure 5. 12). However, Monocle misplaced a few cells, resulting in bimodal expression pattern of Tbr2/Eomes and Sox11 over time, which is counter to the known biology (Figure 5. 12A). In contrast, Waterfall not only placed these cells correctly but also provided on/high and off/low states for a better appreciation of the gene expression state transition (Figure 5. 12B).

2. In vitro myoblast differentiation (in vitro, single cell RNA-seq from Trapnell and Cacchiarelli et al., 2014 Nature Biotechnology)

Trapnell et al. analyzed in vitro human skeletal muscle myoblast (HSMM) differentiation. The HSMM dataset contained 372 single cells: 96 cells were from growth media (GM) before converting to differentiation media (DM). 96 cells were from 24 hours, 96 cells were from 48 hours and 84 cells were from 72 hours after converting to DM. Since Waterfall was developed to reconstruct the in vivo trajectory without any prior information annotated to cells, we merged all conditions, reconstructed the process and compared it with the Monocle predictions with or without the prior information.

Mapping and calculating normalized expression level

We used RSEM to calculate TPM values (Li and Dewey 2011).

Waterfall: Preprocessing

We performed unsupervised clustering and principal component analysis (Figure 5. 13A and C). We could determine that Branch 1 represented single cells from growth media (GM), where cells were highly proliferative with proliferation markers such as PCNA and CDK1 (Figure 5. 13B). On the contrary, Branch 2 represented single cells from a myogenic population in differentiation media (DM) with high myogenic markers such as ENO3 and MYOG (Figure 5. 13B). Since Branch 3 less likely represents myogenic population as opposed to the other two branches, we excluded Branch 3, thereby achieving a single myogenic trajectory from Branch 1 (GM; light purple, light green) to the Center (early stage in DM; red, yellow) and to Branch 2 (late stage in DM; dark orange, light orange) (Figure 5. 13C).

We found that, when given the temporal information, Monocle resulted in a similar trajectory, where Branch 1 largely corresponded to Group 2, Branch 2 largely corresponded to Group 1, and Branch 3 largely corresponded to Branch 3 (Data not shown). Trapnell et al. also excluded the Group 3 because they thought it represented contaminating fibroblasts.

Waterfall: Reconstructing trajectory and pseudotime

We reconstructed the MST by connecting k-means within the PCA plot (Figure 5. 14A) and assigned pseudotime to each individual single cell based on their relative location on the trajectory (Figure 5. 14B).

Waterfall: Gene expression analyses

Most genes showed expected expression patterns, suggesting the reconstructed trajectory reflected the myogenic differentiation process of HSMM (Figure 5. 15).

We sought to compare the performance of Waterfall without prior information, Monocle with prior information, and Monocle without prior information. First, we evaluated the gradual transition of the single cell transcriptomes along the trajectories as in Figure 5. 11 (Figure 5. 16A). Waterfall without prior information performed significantly better than Monocle with or without prior information. Second, we compared the predictions for marker genes from each method. Both

Waterfall and Monocle can predict novel myogenic markers by performing correlation analysis between gene expression and pseudotime progression. In order to test the performance of Waterfall and Monocle, we queried how well these algorithms could find the known myogenic genes (Figure 5. 16B).

3. Embryonic lung development (In vivo, single cell RNA-seq from Treutlein and Brownfield et al., 2014 Nature).

Mapping and calculating normalized expression levels

The dataset contained 199 single cell RNA-seq transcriptomes. We used RSEM to calculate TPM values (Li and Dewey 2011).

Waterfall: Preprocessing

We focused on the E18.5 single cells (79 cells). To select a major trajectory for the continuous process, we performed preprocessing steps of Waterfall.

We performed unsupervised clustering using Pearson correlation, followed by calculating distance matrix from the correlation matrix (Figure 5. 17A). For PC analysis, we followed the approach of Treutlein et al. (Figure 5. 17B) (Treutlein, Brownfield et al. 2014). First, we extracted 36 genes with the highest absolute loadings from each of the four highest contributing principal components. Second, we performed PCA only using these genes. According to Treutlein et al., E18.5 single cell dataset contained unrelated cell types including Clara cells and ciliated cells (Treutlein, Brownfield et al. 2014). Using Scgb1a1, Krt15, and Foxj1, we were able to determine the Clara cell- and ciliated cell-clusters (Figure 5. 17B-C). Furthermore, based on Ager, Pdpn, and Aqp5 expression we determined the differentiating AT1 population. Based on Sftpb and Sftpc expression, we also identified the differentiating AT2 population (Figure 5. 17B-C). Consistent with Treutlein et al., the central population (BP) represented the stem cell population expressing both AT1 and AT2 markers at lower level than the both extremes (Figure 5. 17B-C).

Waterfall: Reconstructing trajectory and pseudotime

As opposed to Treutlein et al. where the authors used group-to-group comparisons to discover key developmental genes (Figure 5. 21), Waterfall allows for reconstruction of the predicted in vivo trajectory and analyses of gene expression at single cell resolution.

We eliminated the Clara cells and ciliated cells to reveal the two trajectories on the PCA plot (Figure 5. 18). We then performed Waterfall trajectory reconstruction for both AT1 and AT2 routes (Figure 5. 19A-B).

AT2 route

We reconstructed AT2 route using MST (Figure 5. 19A, top). We assigned pseudotime of each single cell based on its relative location on the trajectory (Figure 5. 19A, bottom).

AT1 route

We reconstructed the AT1 route using MST (Figure 5. 19B, top). We assigned pseudotime of each single cell based on its relative location on the trajectory (Figure 5. 19B, bottom).

Waterfall: Gene expression analyses

After reconstructing the trajectory, we could query the whole transcriptome profiles throughout the reconstructed in vivo AT1 and AT2 differentiation process (Figure 5. 20). It recapitulated the key discovery of Treutulein et al. with greater temporal resolution (Treutlein, Brownfield et al. 2014): The BP cells were located at the beginning of both pseudotime trajectories and cells in this cluster expressed low level of the both markers. BP cells gained AT1 markers (Figure 5. 20B, left) and lost AT2 markers along the AT1 route (Figure 5. 20A, left). In contrast, the BP cells gained AT2 markers (Figure 5. 20A, right) and lost AT1 markers along AT2 route (Figure 5. 20B, right).

Importantly, we observed that single cell level heterogeneity was resolved by the successful reconstruction of the trajectory. For example, we could appreciate the single cells expressing higher level of AT2 cell marker Clic5 was placed earlier pseudotime points (asterisks in Figure 5. 20A) than the single cells expressing lower level of Clic5 (arrows at the bottom of Clic5 graph in Figure 5. 20A) along the AT1 trajectory, even though Clic5 was not used for reconstructing the trajectory. These results suggest that the pseudotime reflects the whole transcriptomic shift during the developmental process.

In addition, we were able to appreciate the differential timing of gene expression transitions, which was revealed by the bimodal state prediction. Clustering-based differential expression analysis inevitably underestimates developmental heterogeneity (Figure 5. 21). Therefore, single cell resolution reconstruction and bimodal state prediction would be helpful in interpreting the single cell data in higher resolution and allow achieving additional insights.

4. B cell development (In vivo, single cell mass cytometry from Bendall et al. 2014 Cell)

Using their innovative single cell analysis software called Wanderlust, Bendall et al. successfully reconstructed the in vivo trajectory of in vivo B cell development from single cell mass cytometry data. Unfortunately, Wanderlust is not applicable for single cell RNA-seq analysis in its current form. Although Waterfall was developed for single cell RNA-seq, we found that it was also applicable for single cell mass cytometry dataset from Bendall et al.

Waterfall: Preprocessing

We herein present data analysis for the third biological replicate called “Sample C”, although we found that all four biological replicates resulted in consistent and reproducible conclusion. We used the 19 CD markers and HLA-DR, IgD, IgM, Kappa, and Lambda light chain genes as stratifiers for cellular states. Unsupervised clustering and PC analysis resulted in only a single trajectory (data not shown).

Waterfall: Reconstructing trajectory and pseudotime

Using the normalized marker gene expression table, we reconstructed the trajectory (Figure 5. 22, left) and then assigned pseudotime to each single cell data point based on the relative location on the reconstructed trajectory (Figure 5. 22, right).

Waterfall: Gene expression analyses

After reconstructing the trajectory, we performed expression analysis for all the markers in the dataset. Instead of performing HMM, we averaged expression levels within each bin and generated heat maps (Figure 5. 23A) as well as a polynomial regression fitting plot (Figure 5. 23B) to allow direct comparison with the Wanderlust output (Figure 5. 23C)(Bendall, Simonds et al. 2011). Most markers showed expected patterns of expression (Figure 5. 23A-C). Importantly, all four biological replicates generated comparable data without modification in parameters, suggesting the robustness of the Bendall et al. dataset as well as Waterfall (Data not shown).

5. Synthetic datasets

We generated two synthetic datasets. For synthetic dataset I, we allowed gene expression to change gradually over the sequence of single cell transition. For synthetic dataset II, we generated a more realistic condition that contains a branch coming out from a primary trajectory. We also introduced random stochasticity for dataset II.

Waterfall: Preprocessing

For synthetic dataset I, the variance contributions were distributed to four principal components (Figure 5. 24D). We demonstrated the gene expression level by the size of each single cell data point proportional to the gene expression level (Figure 5. 24D).

For synthetic dataset II, we introduced a branch as well as expression stochasticity (Figure 5. 24A, left and middle). Single cells originating from this branch were mixed with the single cells originating from the major trajectory of interest on the PCA plot (note the single cells with an asterisk at Figure 5. 25A, right). However, hierarchical clustering precisely determined the branch-derived single cells (Figure 5. 25B-C), so that we could exclude this branch from further analyses.

Waterfall: Reconstructing trajectory and pseudotime

For synthetic dataset I, the color-coded clusters lined up sequentially throughout the trajectory. K-mean clustering and MST successfully approximated the average trajectory (Figure 5. 26A), thereby generating a near linear graph without significant tangential offset (Figure 5. 26A).

For synthetic dataset II, after the branch was removed the major trajectory was successfully approximated by k-means and MST (Figure 5. 27A). When it was linearized and flattened as shown in the Figure 5. 27B, we were able to appreciate the stochasticity that we introduced forming the tangential distribution of the pseudotime progression (Figure 5. 27B).

Waterfall: Gene expression analyses

We checked a few gene expression profiles in the synthetic dataset I. As expected, the gene expression profile underwent smooth transition throughout the trajectory suggesting the reconstruction was precise (Figure 5. 28). Of note, the HMM was able to predict the transition points of each gene. Under naturally occurring conditions, these predictions could be validated by independent biological experiments.

Table 5. 1. List of datasets to which Waterfall was applied.

Dataset	Biological Process	Single Cell Modality	Bioinformatic Algorithms	Figures
Current study	*In vivo* adult neurogenesis	RNA-seq	Waterfall & Monocle	Figures 5. 1 –11
Trapnell et al., 2014	*In vitro* myogenesis	RNA-seq	Waterfall & Monocle	Figures 5 .12 – 15
Treutlein et al., 2014	*In vivo* lung development	RNA-seq	Waterfall	Figures 5. 16 – 20
Bendall et al., 2014	*In vivo* B cell development	Mass-cytometry	Waterfall	Figures 5. 21 – 22
Current study	Synthetic dataset I & II	–	Waterfall	Figures 5. 23 – 27

Table 5. 2 . Comparison between static and dynamic heterogeneity.

	Static/cell-type heterogeneity	**Dynamic/temporal heterogeneity**
Goals	Discovering novel cell types from a seemingly homogeneous population	Understanding the continuous biological process of one cell type
Dataset	Multiple cell types	One cell type at different temporal stages
Analytical approach	Clustering-based comparison (analogous to population RNA-seq analysis after FACS sorting)	Ordering single cell transcriptomes, based on the transcriptomic similarity (Waterfall or Monocle)
Examples	Unappreciated heterogeneity within sensory neurons in dorsal root ganglia (Usoskin, Furlan et al. 2015)	Molecular cascades throughout adult neurogenesis (Current study)

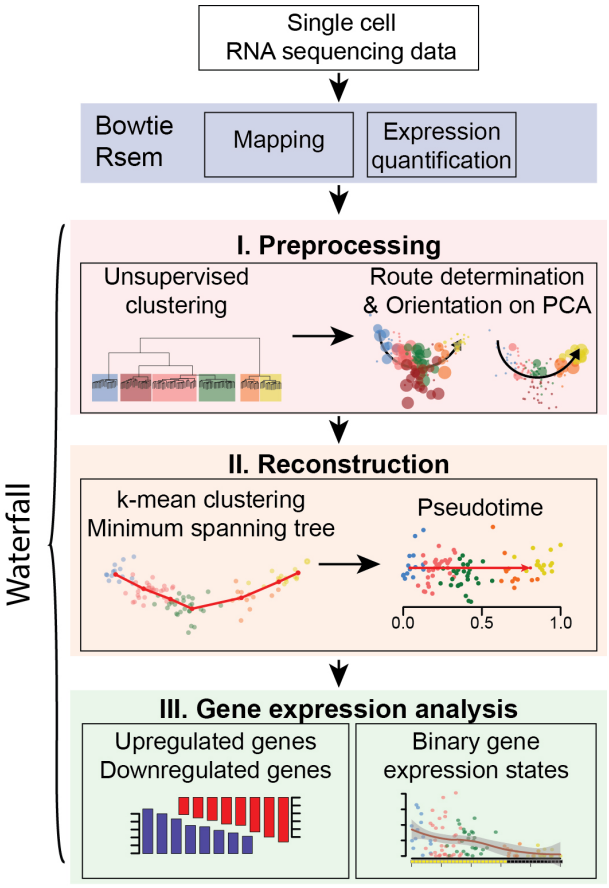

Figure 5. 1. Waterfall pipeline consists of three key steps of Pre-processing, Reconstruction, and Gene expression analysis.

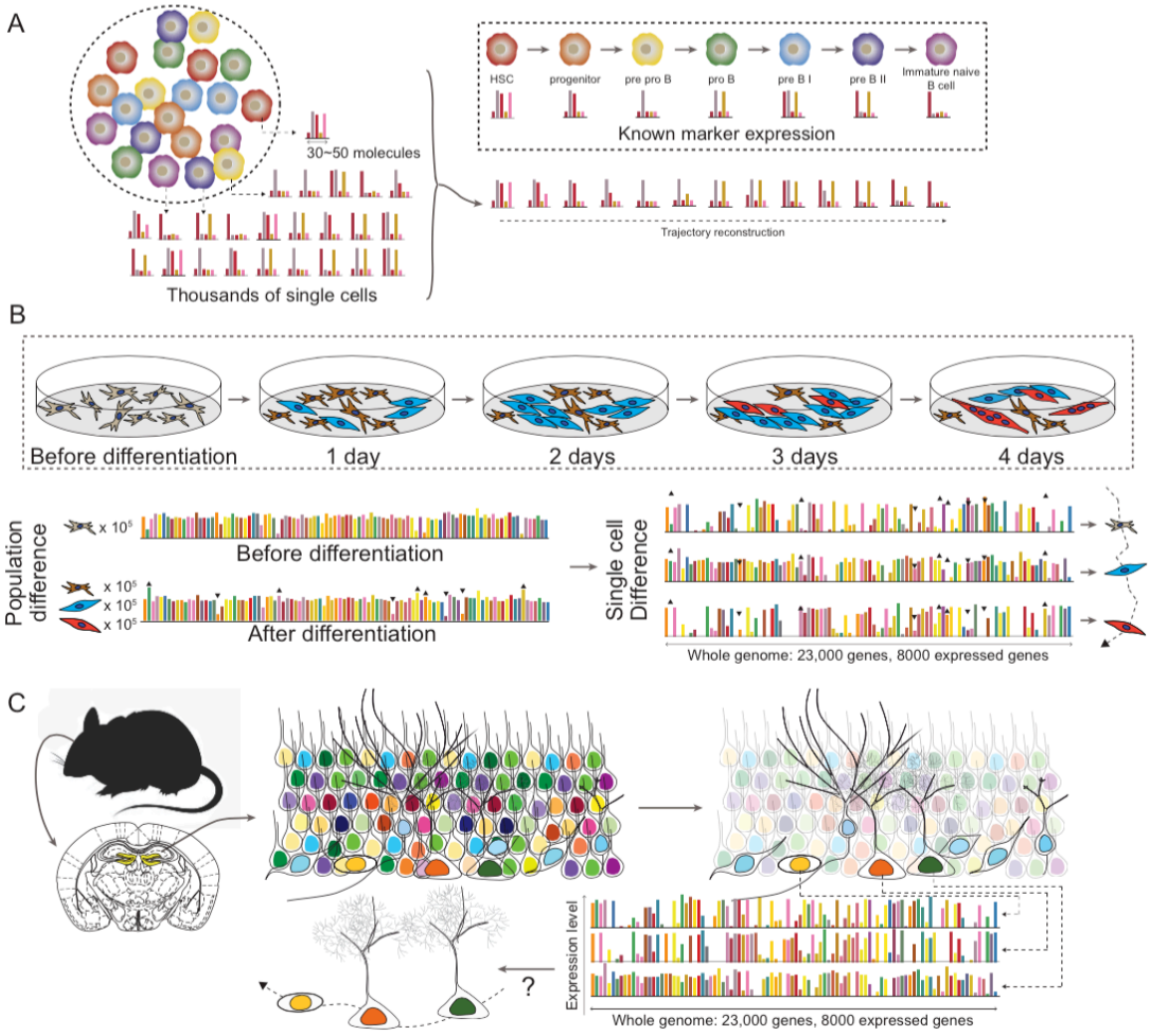

Figure 5. 2. Schematic comparison of the complexity between single cell qPCR or mass cytometry.

Single cell qPCR or mass cytometry generates data with 30~50 molecular markers for thousands of single cells (A). Using the known marker genes as stratifiers, we can order the single cells (A). For in vitro conditions, we can discover developmentally important marker genes by performing differential analysis using temporal information of single cells at population level (B). For in vivo conditions, however, we have a completely mixed population of differential developmental points (C). Thus, reconstructing the trajectory from the in vivo single cell is a significant challenge.

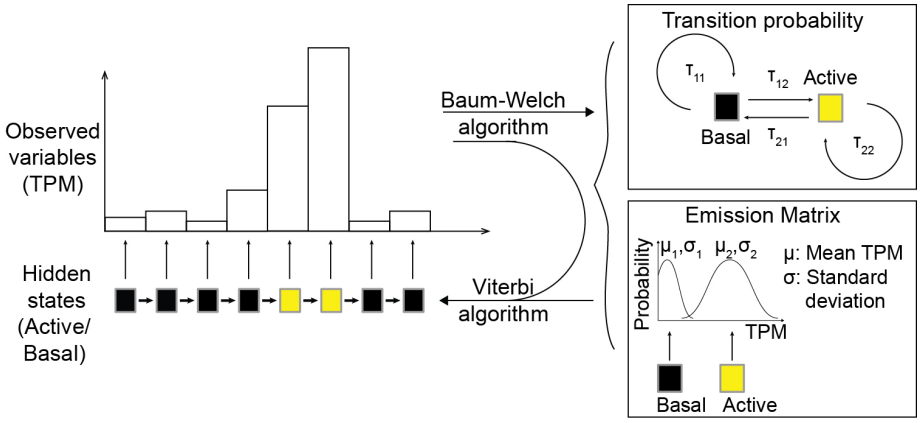

Figure 5. 3. Hidden Markov model (HMM). We predicted underlying states from gene expression (TPM) over pseudotime progression.

Baum-Welch algorithm predicts the most likely transition probability and emission matrix from observed variables (TPM). Viterbi algorithm uses observed variables (TPM) along with output from the Baum-Welch algorithm to predict hidden On/High and Off/Low states.

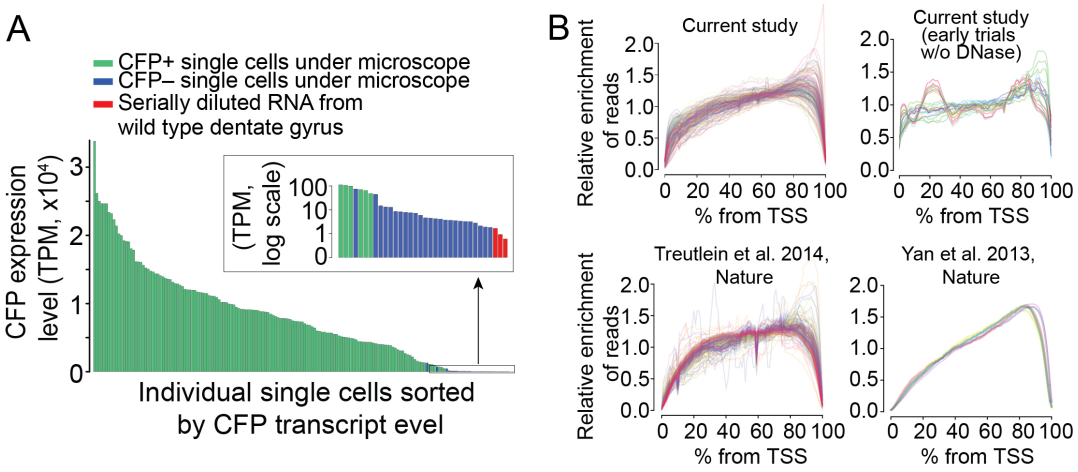

Figure 5. 4. Precision of CFPnuc+ single cell isolation (A) and minimal cDNA amplification related 3' bias (B).
Dnase treatment significantly augmented the quality of the sequencing (B).

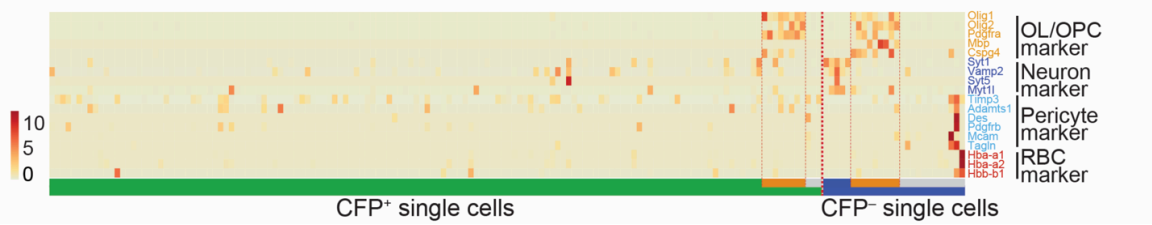

Figure 5. 5. Marker based cell type screening. Single cells expressing known markers for pericytes, oligodendrocyte (OL), oligodendrocyte progenitor cells (OPC) were excluded for the further analyses.

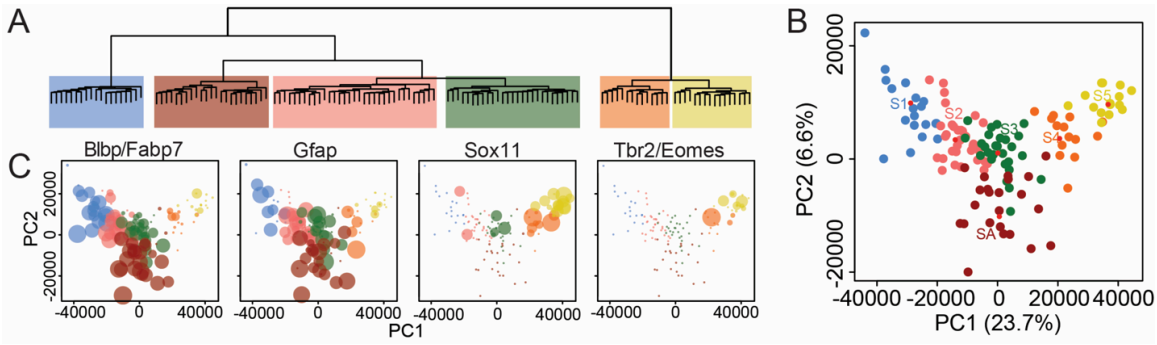

Figure 5. 6. Preprocessing step of Waterfall - Unsupervised clustering (A), principal component analysis (B) and marker expression patterns (C) to help orient of data.

Color of each data point follows the same color code. Size of the each data point in C represents normalized expression level.

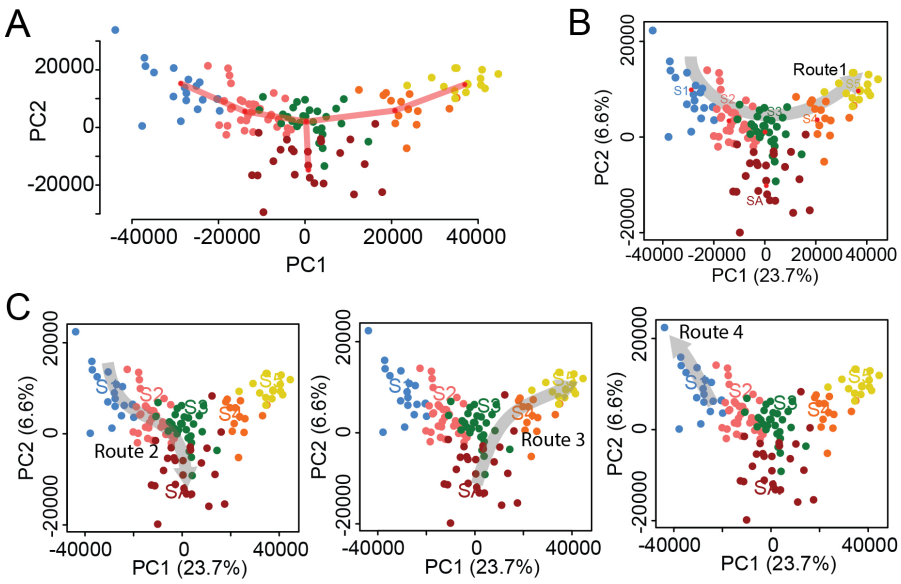

Figure 5. 7. Potential trajectories of adult neurogenesis dataset- route definition using minimum spanning tree with k-means clustering (A), the major neurogenic trajectory (B) and alternative routes to consider (C).

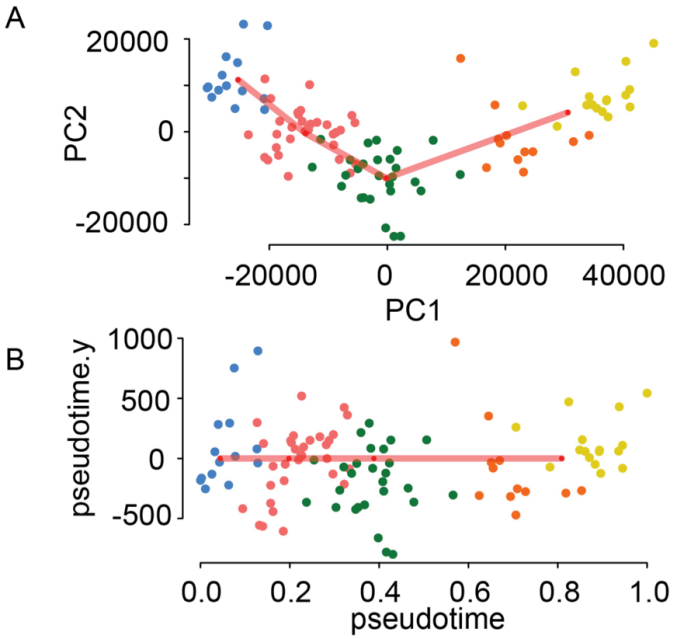

Figure 5. 8. Trajectory reconstruction step of Waterfall. The trajectory was built using MST connecting the five vertices from k-means clustering (A).

Pseudotime of each data points was assigned by its relative location when projected on to the reconstructed trajectory (B). The orientation was determined by Figure 5. 6C.

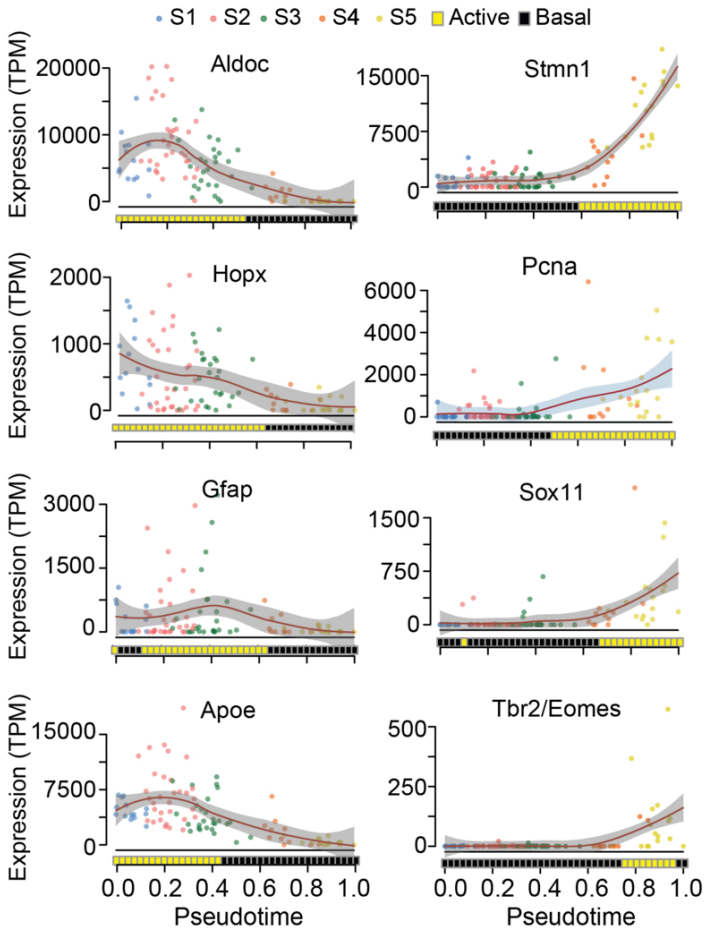

Figure 5. 9. Representative gene expression dynamics predicted by Waterfall. Shown are scatterplots of single cell gene expression levels.

We demonstrated local polynomial regression fitting plot with 95% confidence interval (red linear graphs and gray shades). HMM-predicted transcriptional states throughout pseudotime progression were demonstrated as block graphs at the bottom of each plot with on/high states as yellow blocks and off/low states as black blocks.

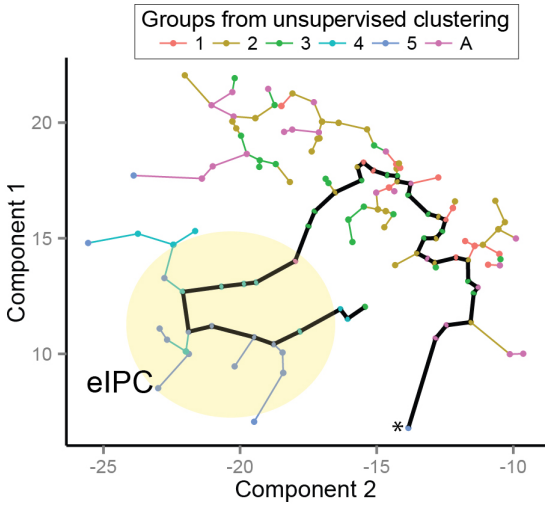

Figure 5. 10. Monocle-derived trajectory using top 1000 highly expressed genes.

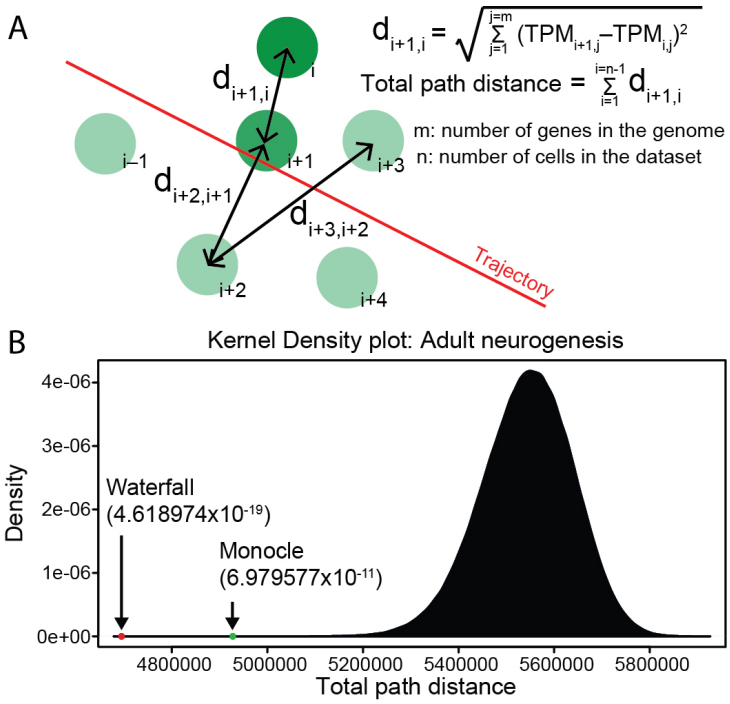

Figure 5. 11. The quantitative test for gradual transcriptomic transition of single cell trajectories.

(A) The distance between a pair of single cells is defined by the square root of the sum of square of each gene expression difference. Total path distance is defined by the sum of the distances between neighboring single cell pairs along the trajectory from the beginning to the end of the trajectory. (B) Kernel density plot represents the distribution of total path length from random ordering of single cells. Total path lengths of Waterfall- and Monocle-derived pseudotime order of single cells were marked by arrows. The cumulative probabilities for each predicted path were 4.6 $x10^{-19}$ for Waterfall and 7.0 $x10^{-11}$ for Monocle.

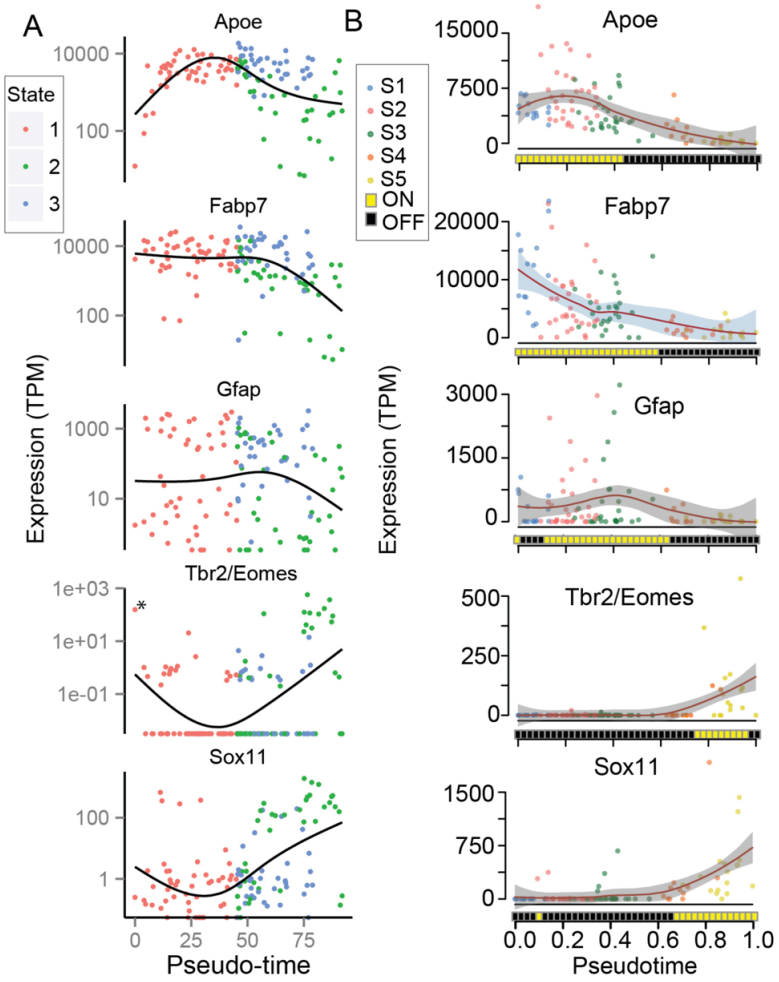

Figure 5. 12. Expression profiles of key neurogenic genes from Monocle (A) and Waterfall (B).

Note that Monocle represents the data as a logarithmic scale whereas Waterfall represents the data as a linear scale. The single cell with an asterisk at the Monocle predicted Tbr2/Eomes profile indicates one of misplaced cells, which was also marked at Figure 5. 10.

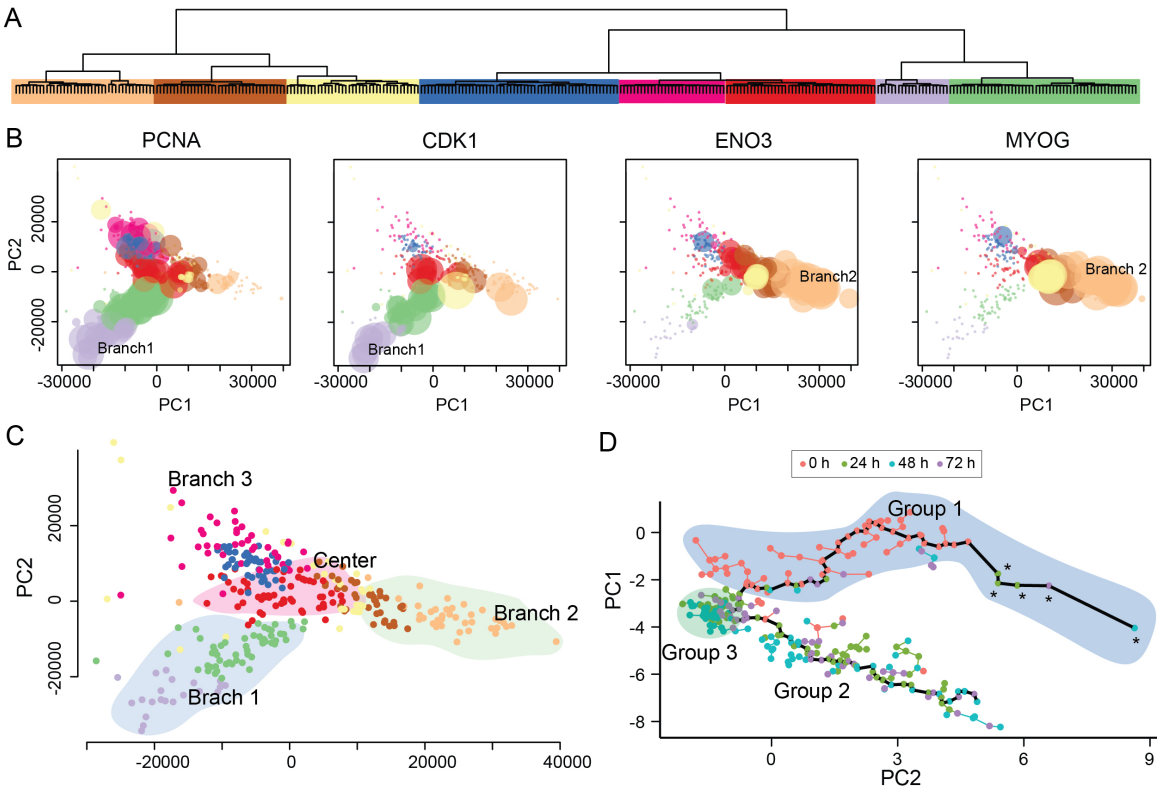

Figure 5. 13. Preprocessing steps of Waterfall: Unsupervised clustering (A), marker gene expression profiles (B), Waterfall trajectory determination without prior temporal information (C); and Monocle trajectory with prior temporal information as a comparison (D).

The color in B and C follows the color code appearing in A. The size of each data point at B represents normalized expression value.

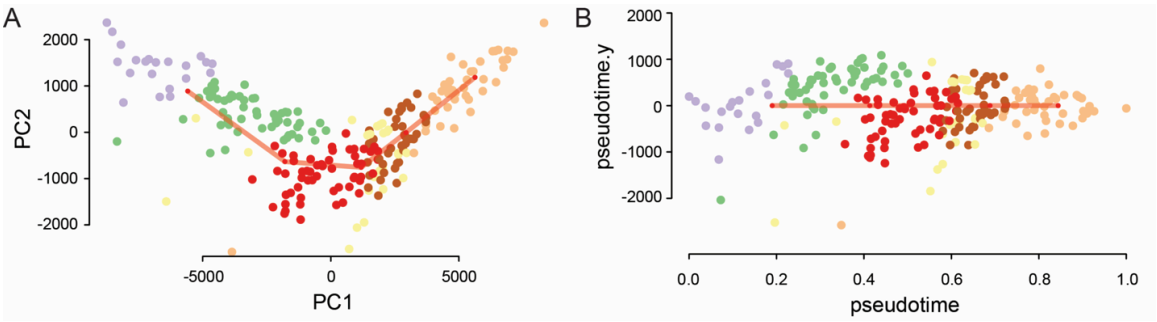

Figure 5. 14. Pseudotime reconstruction of selected route of in vitro HSMM differentiation. MST connected k-means to generate trajectory (A).

Pseudotime was determined based on the relative location of single cells when projected onto the reconstructed trajectory (B). The color of each data point follows the color code at Figure 5. 13 A.

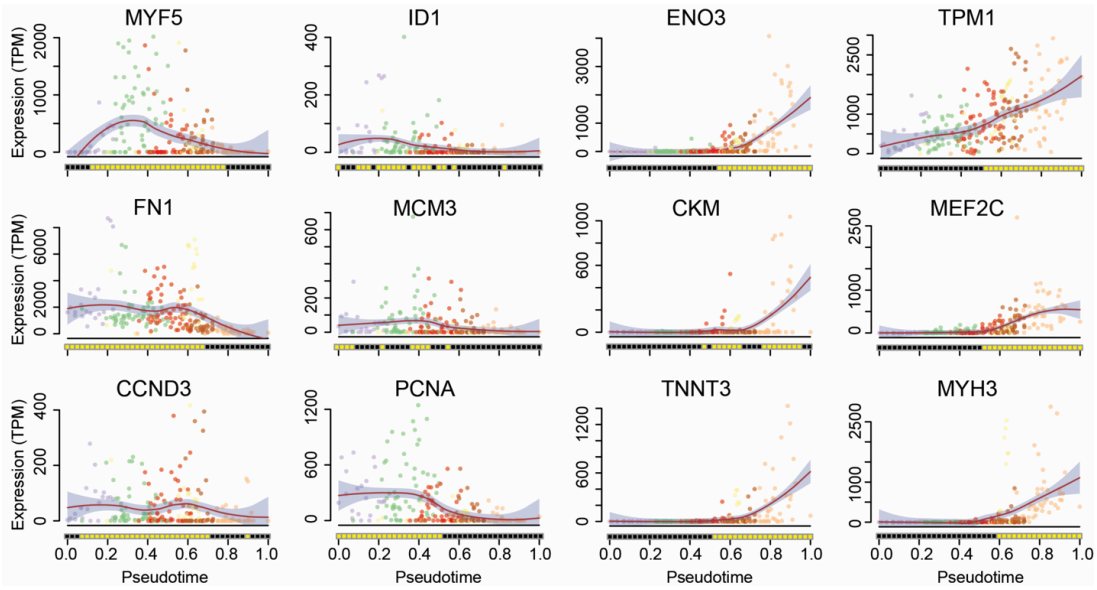

Figure 5. 15. Gene expression dynamics over the reconstructed HSMM differentiation process.

Shown are the local polynomial regression fitting plots with 95% confidence interval (red linear graphs and gray shades), and HMM-predicted transcriptional states (block graphs at the bottom; on/high-yellow, off/low-black). Proliferative marker genes or primitive muscle developmental genes were highly expressed at earlier pseudotime points (left two columns) whereas skeletal muscle markers were highly expressed at later pseudotime points (right two columns).

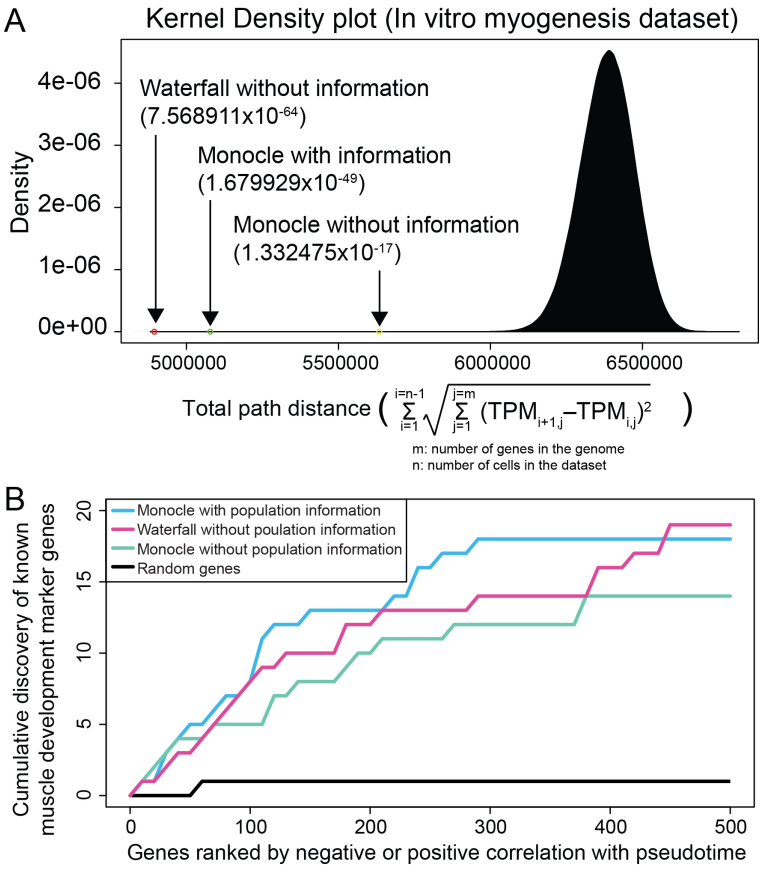

Figure 5. 16. Comparison between Waterfall and Monocle for in vitro myogenesis dataset.

(A) Kernel density plot represents the distribution of the total path length of random order of single cells. Arrows indicate the total path length of pseudotime order of single cells from Waterfall without prior information, Monocle with prior information and Monocle without prior information. The cumulative probabilities for each predicted path were 7.6×10^{-64} for Waterfall without prior information, 1.7×10^{-49} for Monocle with prior information and 1.3×10^{-17} for Monocle without prior information. (B) Discovery of myogenic genes using different methods of trajectory building. We counted the number of known myogenic genes within the top 500 genes with highest positive or negative absolute correlation with pseudotime progression predicted from (1) Monocle with prior information (light blue), (2) Waterfall without prior information (pink), (3) Monocle without prior information (light green) and (4) randomly picked genes (dark gray). List of markers for muscle development was from Trapnell et al.(Trapnell, Cacchiarelli et al. 2014) and http://www.rndsystems.com/molecule_group.aspx?g=821.

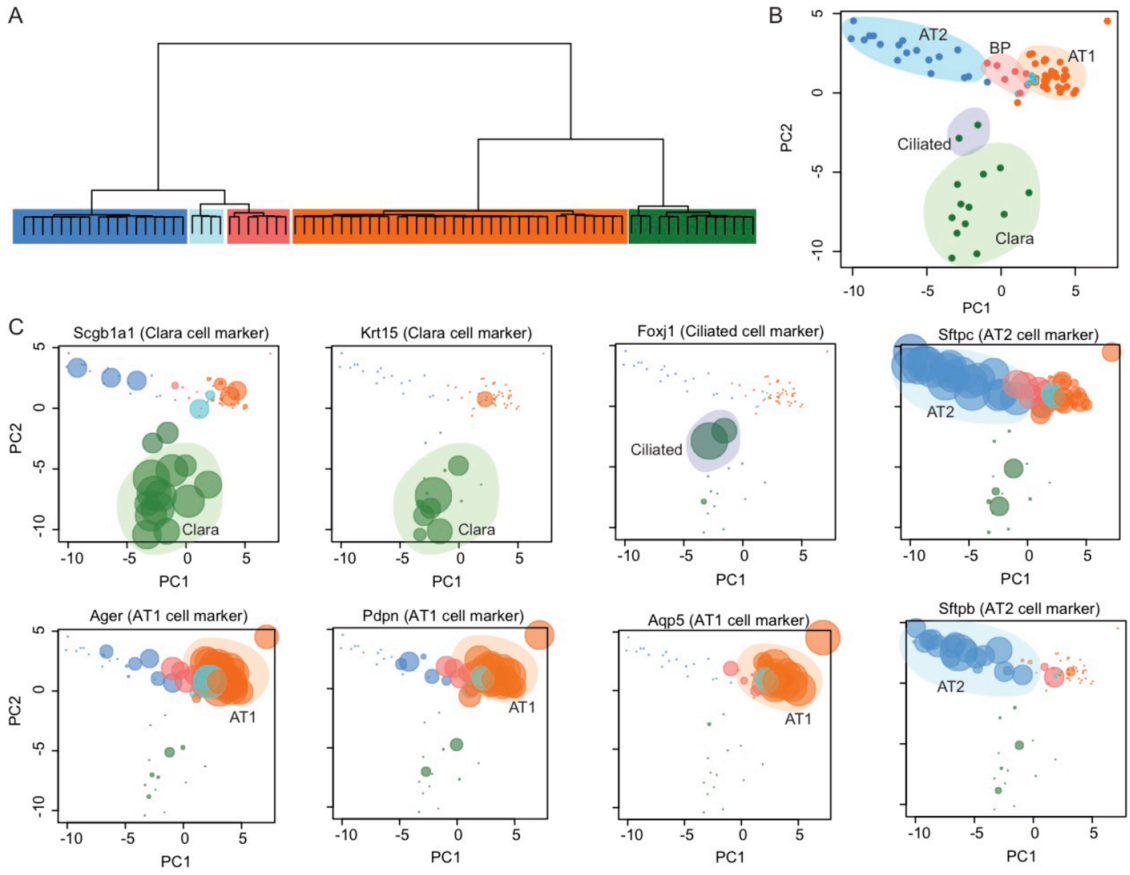

Figure 5. 17. Preprocessing step of Waterfall for embryonic lung development: including unbiased clustering (A), PCA plot using selected genes from the four major PC axes (B), and marker gene expression profiles to orient and define a single route of interest (C).

Size of each data point in C represents normalized gene expression and the color follows the color codes in A.

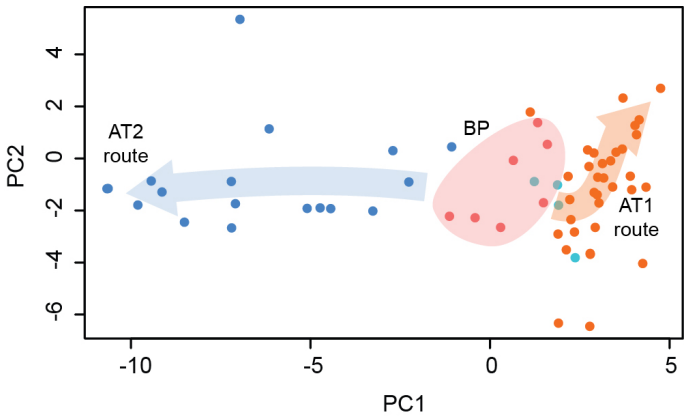

Figure 5. 18. PCA plot after removing Clara/ciliated cell group.

AT1 route and AT2 route are shown with orange and blue arrows, respectively.

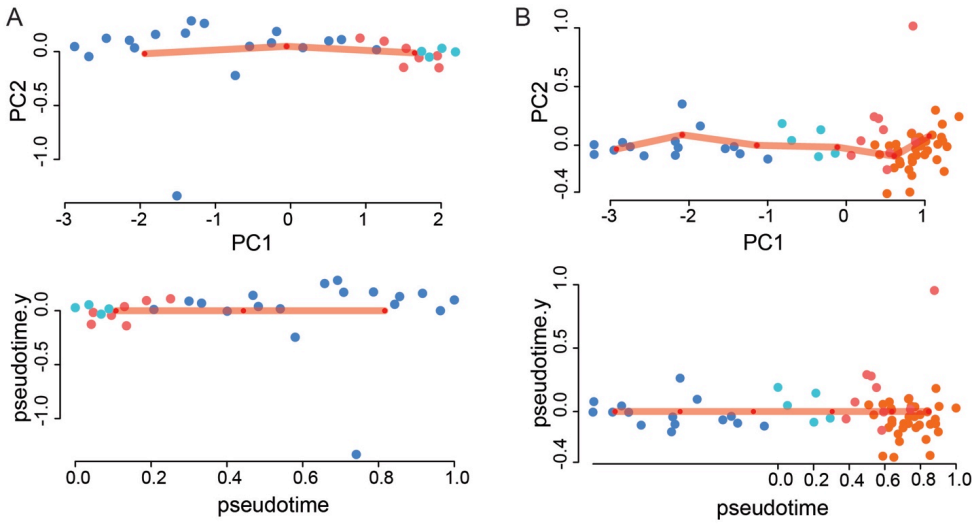

Figure 5. 19. Reconstructing continuous trajectory and assigning pseudotime for AT2 (A) and AT1 differentiation process (B).

The vertices from the k-means clustering were connected by MST (A, top; B, top). Pseudotime was determined by the relative location of single cells when projected on to the trajectory (A, bottom; B, bottom). The color of each data point follows the color code appearing in Figure 2. 3

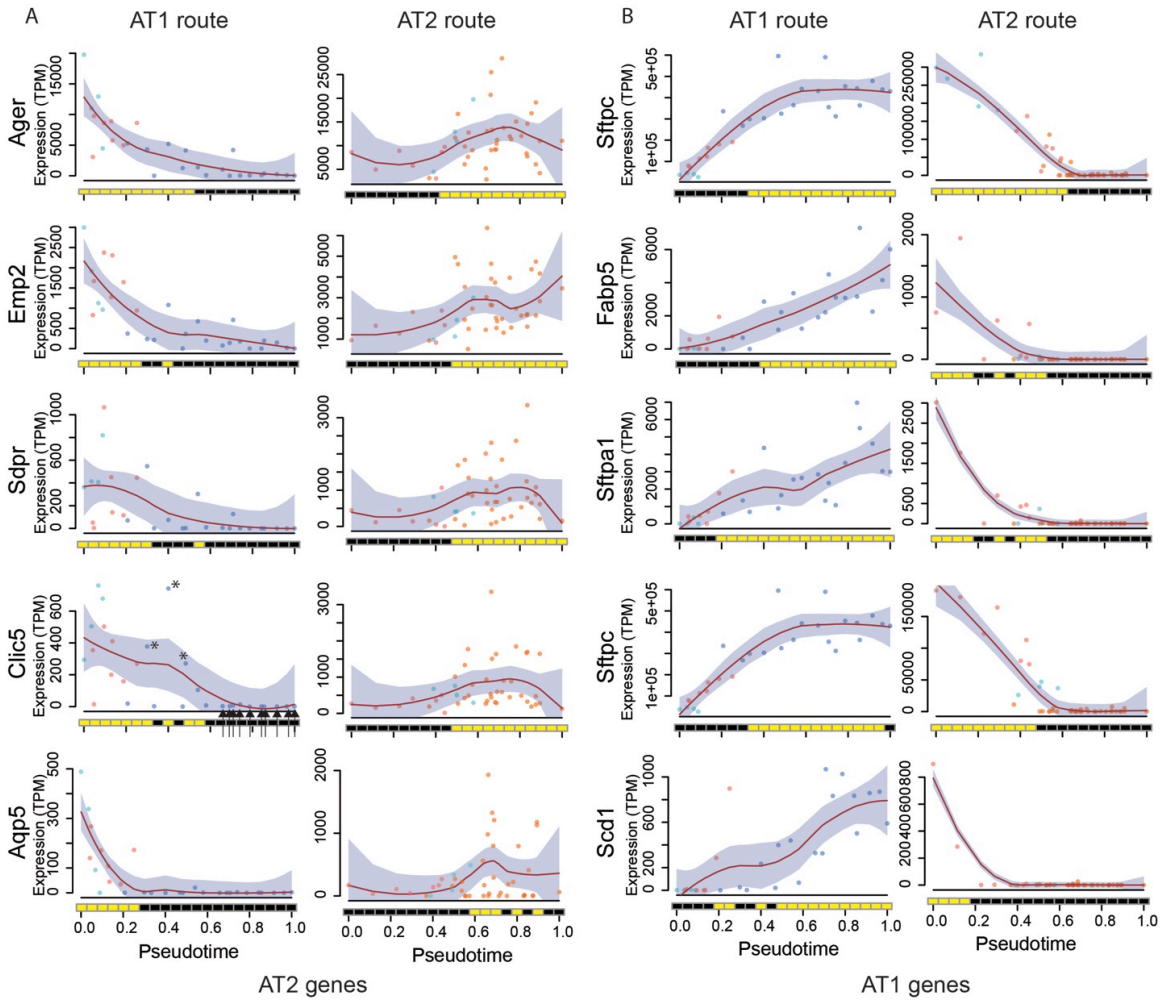

Figure 5. 20. Waterfall gene expression prediction profile for AT2 markers (A) and AT1 markers (B).

We showed the local polynomial regression fitting plot with 95% confidence interval (red linear graphs and gray shades), and HMM-predicted transcriptional states (block graphs at the bottom; on/high-yellow, off/low-black). BP cells gained AT1 markers (B, left) and lost AT2 markers through AT1 route (A, left). In contrast, BP cells gained AT2 markers (A, right) and lost AT1 markers through AT2 route (B, right). Even though single cells indicated by the arrows and single cells indicated by asterisks at Clic5 expression profile were classified as a same group by unsupervised clustering, Waterfall successfully placed them accordingly, further highlighting the greater resolution of Waterfall. Of note, Clic5 was not one of the genes that was used for trajectory reconstruction.

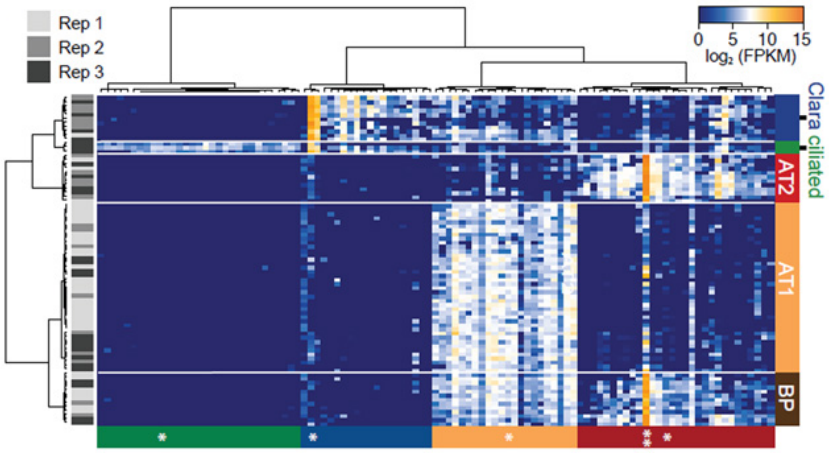

Figure 5. 21. Cluster based differential analysis adopted from Figure 2A of Treutlein et al. (Treutlein et al., 2014).

Rows represent clustered single cells and columns represent clustered genes.

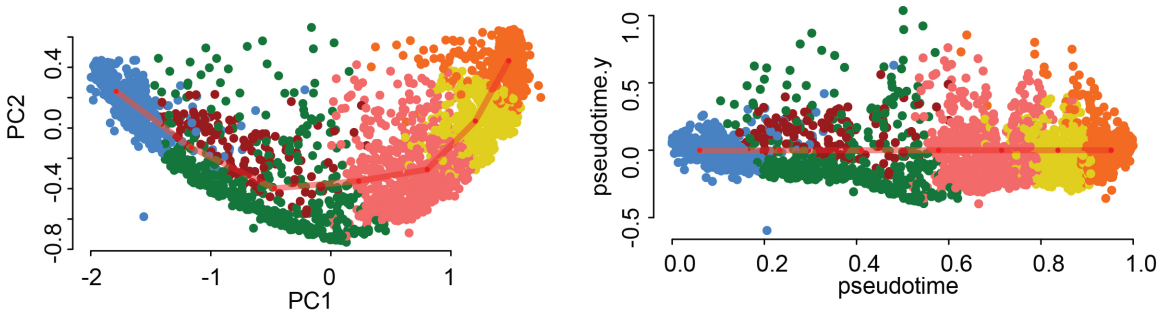

Figure 5. 22. Waterfall trajectory reconstruction for the "Sample C" of Mass cytometry B cell development dataset. Vertices from the k-means clustering were connected by MST (left).

Pseudotime of each data points was assigned by the relative location when projected on to the reconstructed trajectory (right). The color of each data point follows the color codes from the unsupervised clustering (data not shown).

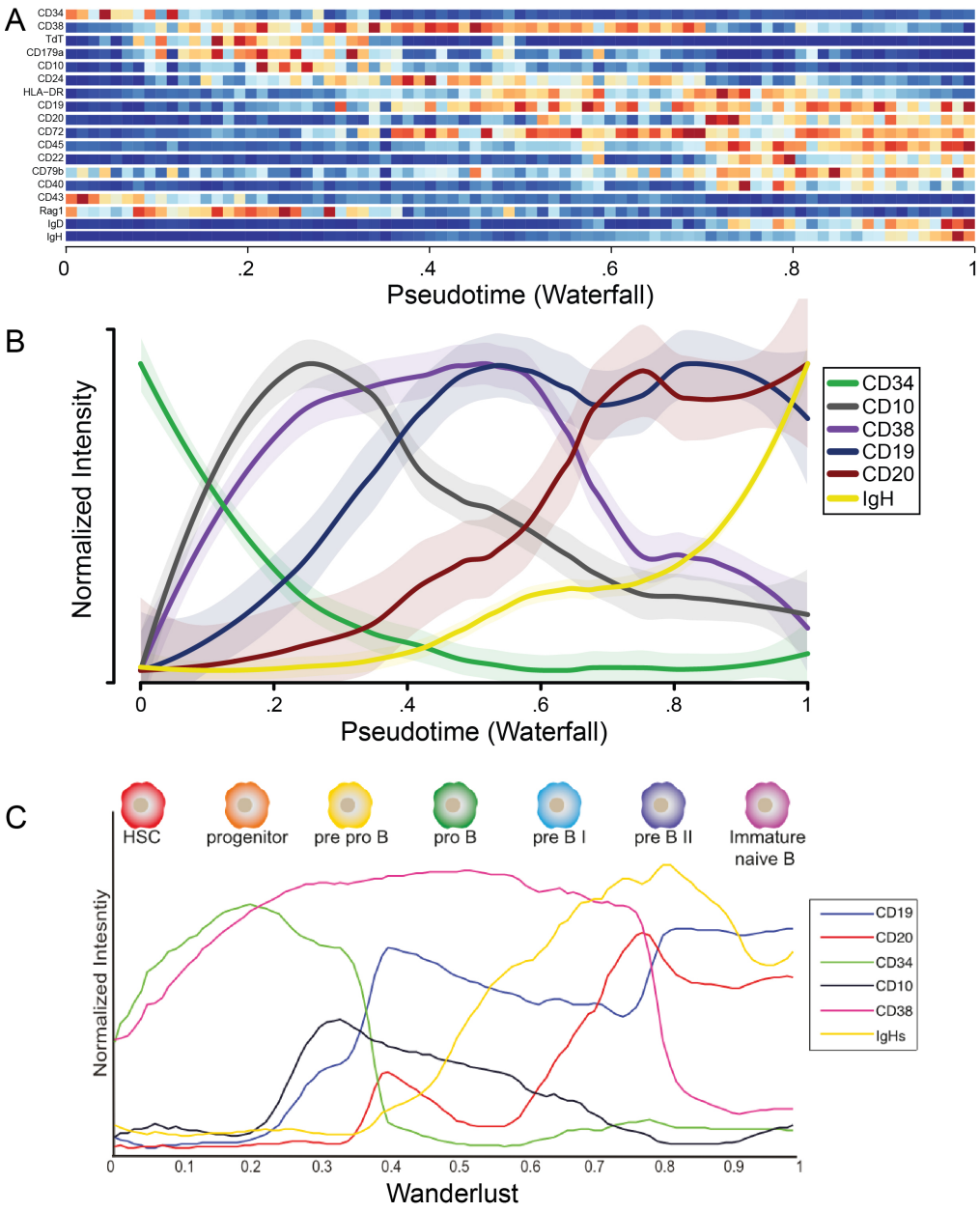

Figure 5. 23. Waterfall predicted gene expression profiles along the B cell developmental trajectory from the "Sample B" (A and B) and the Wanderlust predicted profile (C).

(A) The reconstructed trajectory by Waterfall was divided into 80 bins and the mass cytometry-measured antigen abundance was averaged within each bin. (B) We selected the key marker antigens for B cell development from Waterfall analysis. We showed the local polynomial regression fitting plot with 95% confidence interval (shaded region). (C) Key marker antigen profiles for B cell development, taken from Bendall et al. (Bendall, Davis et al. 2014).

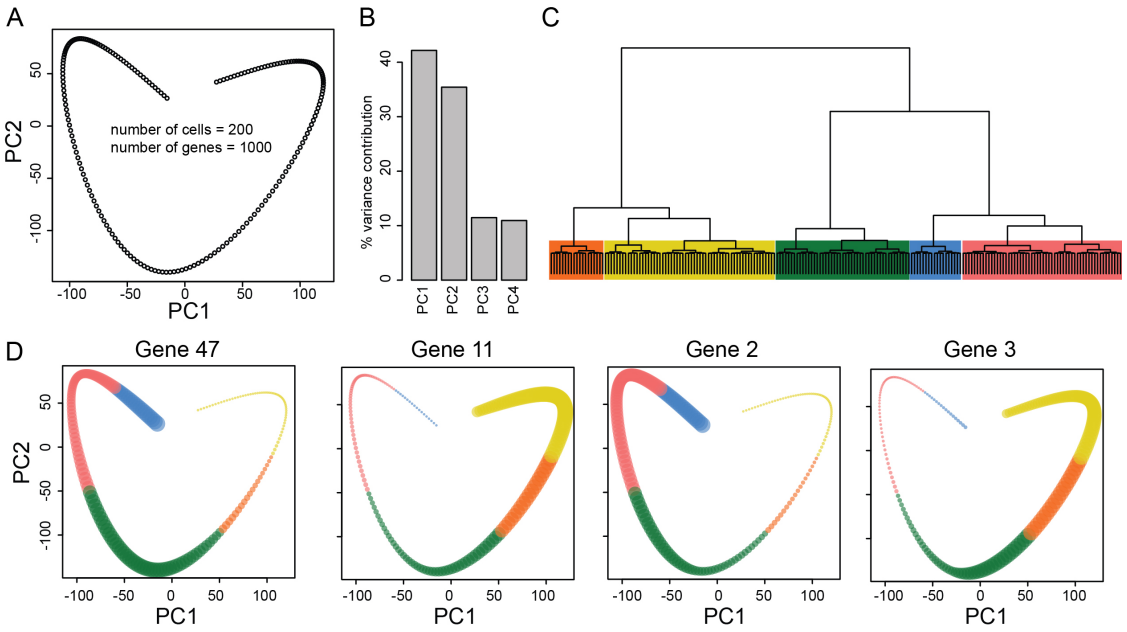

Figure 5. 24. Preprocessing of Waterfall analysis for synthetic dataset I.

The dataset consisted 1000 hypothetical genes gradually changing throughout preset order of 200 single cells, thereby generating a linear trajectory on the PCA plot (A). Four principal components could explain the variance (B). Although unsupervised clustering resulted in clusters, the clusters were not robust due to the continuity nature of the dataset (C). Examples of hypothetical gene expression were shown to mimic the preprocessing step of Waterfall (D). The size of the each data point is proportional to the gene expression level and the color of the each data point follows the color code in C.

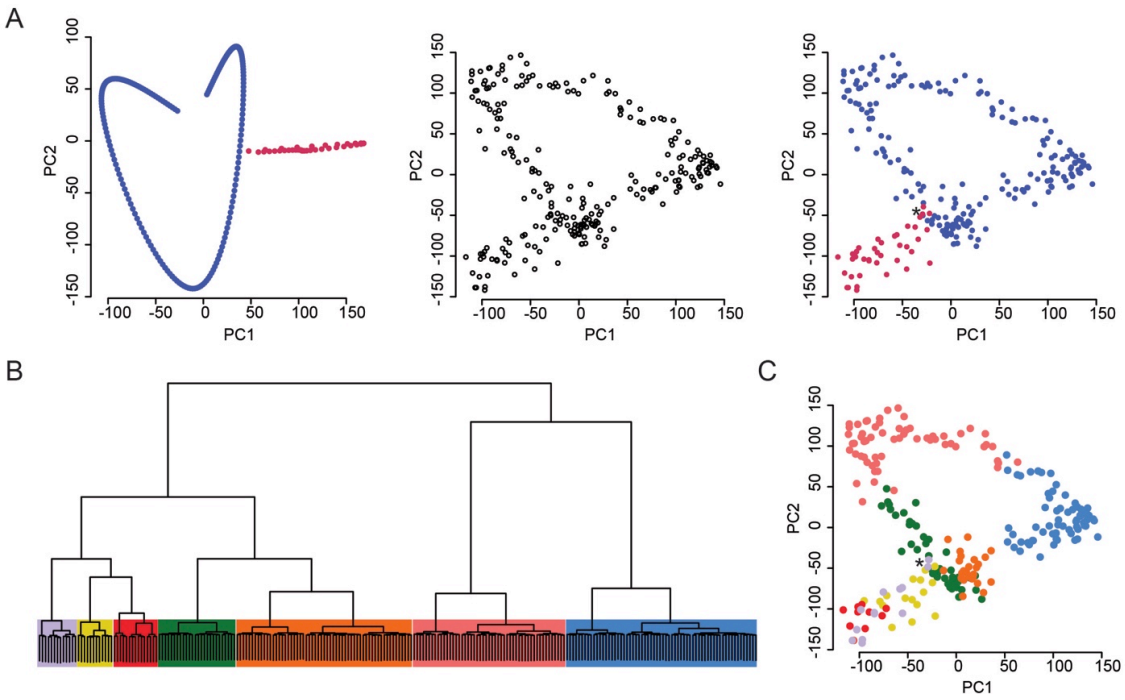

Figure 5. 25. Preprocessing of Waterfal analysis for the synthetic dataset II.

The linear dataset with a branch (A, left) was introduced with stochasticity (A, middle). It was ambiguous where the single cells close to the branching point belong to (A, right, asterisk), but unsupervised clustering (B) successfully determined them (C, asterisk). We cut out single cells that belong to the branch for the further analysis, which is a simulation for preprocessing for the real dataset appearing at Figure 5. 7 and Figure 5. 13.

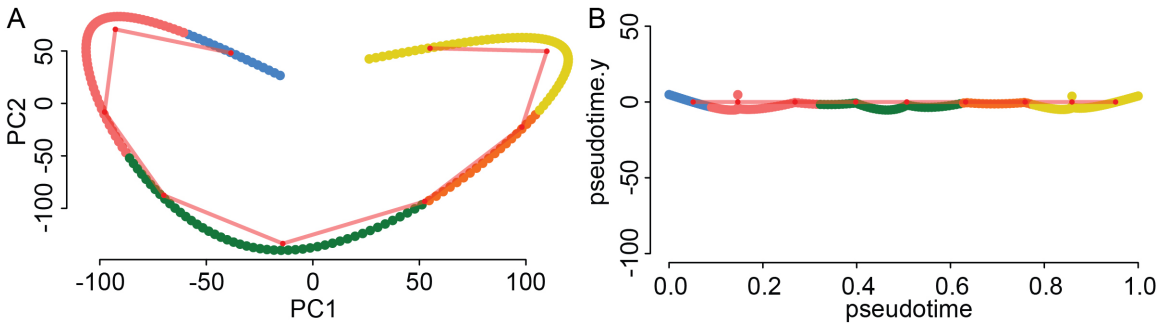

Figure 5. 26. Waterfall reconstructed trajectory and pseudotime assignment for synthetic dataset I.

Vertices from the k-means clustering were connected by MST (A). Pseudotime of each data point was assigned by the relative location when projected on to the reconstructed trajectory (B). The color of each data point follows the color codes appearing in Figure 5. 24C.

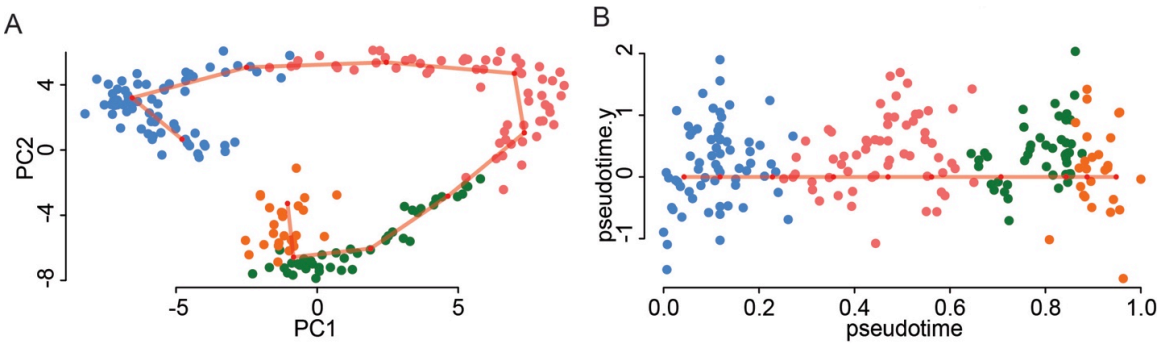

Figure 5. 27. Waterfall reconstructed trajectory and pseudotime assignment for synthetic dataset II.

Vertices from the k-means clustering were connected by MST (A). Pseudotime of each data points was assigned by the relative location when projected on to the reconstructed trajectory (B). The color of each data point follows the color code in Figure 5. 25B.

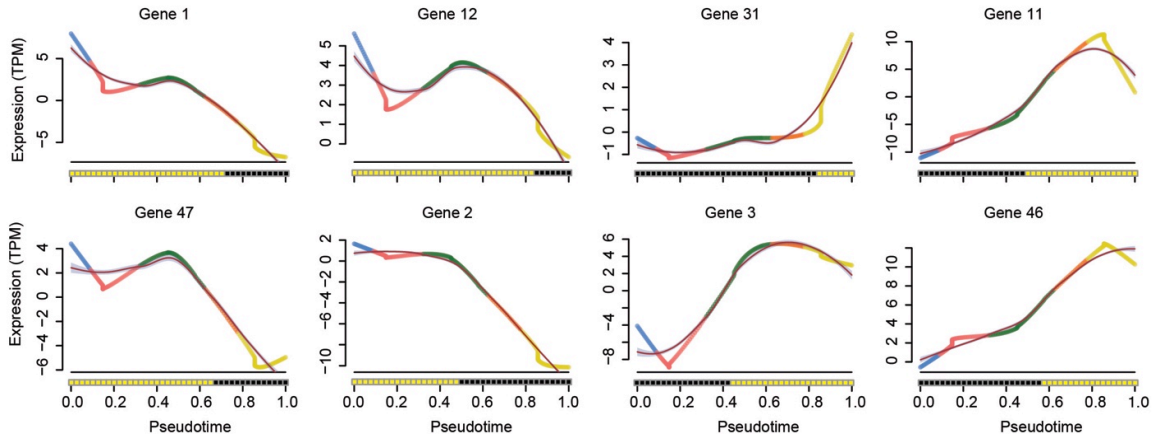

Figure 5. 28. Gene expression profiles for synthetic dataset I, predicted by Waterfall.

The color of each data point follows the color code in Figure 5. 24C. We showed the local polynomial regression fitting plot (red linear graphs), and HMM-predicted transcriptional states (block graphs at the bottom; on/high-yellow, off/low-black).

Chapter 6. Experimental procedures[6]

[6] This chapter is based on Shin, J., et al. (2015). "Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis." Cell Stem Cell **17**(3): 360-372.

Preparation of Individual Cells from Adult Mouse Dentate Gyrus

Homozygous transgenic mice expressing nuclear localized CFP (CFP^{nuc}) driven by the Nestin promoter (Encinas et al., 2006) were used for all single-cell RNA-seq experiments. Mice were euthanized by cervical dislocation, and brains were immediately immersed into ice cold Dulbecco's Phosphate-Buffered Saline (DPBS, Corning). All procedures were performed with approved protocols in accordance with institutional animal guidelines.

The dissected dentate gyrus was incubated in Hibernate A (BrainBits) containing papain (100 U; Sigma) and RNase-free DNase I (100 units; NEB) at 37˚C for 18 minutes with intermittent flicking. The tissue was triturated into individual cell suspension by 1 ml pipette (Denville Scientific). Enzymes and cellular debris were removed with multiple rounds (~4-5 times) of mild centrifugation at 200g and washing with Hibernate A minus Ca^{2+} and Mg^{2+} (BrainBits). The individual cell suspension was plated onto a glass bottom plate (MatTek) and picked up using glass pipettes (World Precision Instruments) under a fluorescent microscope. The glass tip was broken into the bottom of each PCR tube containing water (2.4 μl) with RNase-free DNase I (0.2 μl; NEB) and Murine origin RNase inhibitor (0.25 μl; NEB). Importantly, the addition of DNase I significantly improved the quality of the data by removing contamination from the random amplification of genomic DNA (See Chapter 5. Single cell analysis with Waterfall and Figure 5. 4B).

Library Preparation and Sequencing

cDNA amplification followed the previously published SMART protocol (Ramskold et al., 2012). Briefly, the DNase I was first inactivated by increasing the temperature (75˚C for 10 minutes) and samples were then stored on ice. Custom designed 2A oligo 1 μl (12 μM, Integrated DNA Technologies, sequence shown in Figure 2. 2A) was added and annealed to the polyadenylated RNA by increasing temperature (75˚C for 3 minutes) and quenching on ice. A mixture of 2 μl Superscript II First-Strand Buffer (5X, Invitrogen), 1 μl custom designed TS oligo (12 μM, Integrated DNA Technologies, Figure 2. 2A), 0.3 μl $MgCl_2$ (200 mM, Sigma), 0.5 μl RNase inhibitor (Neb), 1 μl dNTP (10 mM each, Thermo), 0.25 µl DTT (100 mM, Invitrogen), and 1 μl Superscript II (200 U/µl, Invitrogen), were incubated at 42˚C for 90 minutes, followed by enzyme inactivation at 75˚C for 10

minutes. A mixture of 29 μl Water, 5 μl Advantage2 taq polymerase buffer, 2 μl dNTP (10 mM each, Thermo), 2 μl custom designed PCR primer (12 μM, Integrated DNA Technologies, Figure 2. 2A), 2 μl Advantage2 taq polymerase was directly added to the reverse transcription product and the amplification was performed for 19 cycles. The amplification product was purified using Ampure XP beads (Beckman-Coulter). Library preparation was performed using Ovation Ultralow library systems (Nugen inc). Libraries were multiplexed and sequenced using Illumina Hiseq 2500 (Illumina Inc) (Table 2. 1).

Bioinformatic Analyses

Mapping and calculating gene expression levels: Raw reads were first trimmed for their Illumina adapter sequences using Trimmomatic (Bolger et al., 2014). Custom R codes were used to further trim the 5' TS oligo sequences, and the 3' primer sequences. RSEM (Li and Dewey, 2011) was used to map and calculate gene expression levels represented as transcripts per million (TPM). The reference genome was modified to include chrC, which contained the sequence of part of Nestin enhancer followed by eCFP transcripts, and reconstructed from sequencing reads. We used following parameters: rsem-calculate-expression -p 12--fragment-length-mean 500 $input.fastq $rsem_ref $cell_id. For most of the downstream analyses, we used a table with single cells at the columns and the genes at the rows. For the purpose of visualization (Figure 2. 1C), reads were separately processed using bowtie (Langmead and Salzberg, 2012) and tophat (Trapnell et al., 2009) with the following option: tophat2 -p 8 -N 2 -o $cell_id -g 1 -G $gtf --transcriptome-index $transcriptomeindex $bowtie2index $input.fastq.

Waterfall 1–pre-processing: Waterfall input is an expression matrix from RSEM after eliminating outliers (Figure 2. 2C). Unsupervised clustering was performed using a distance matrix based on Pearson correlation between each pair of single cells (Figure 2. 3A). We defined the neurogenic trajectory on the PCA plot and determined the direction using known markers such as Sox11, Tbr2, Blbp, and Gfap. Expression levels represented on the PCA plot at Figure 2. 4B were based on custom normalization using sine normalization.

Waterfall 2 - Building an in vivo trajectory: We used custom R codes to determine pseudotime for each single cell on the trajectory (See Chapter 5. Single cell analysis with Waterfall). Briefly, we performed parametric PCA, and extracted k-means from the distribution of single-cell transcriptomes (Figure 2. 6A). We generate an unbiased trajectory by connecting k-means centers using a minimum spanning tree (MST) algorithm (Paradis et al., 2004). First, we set zero for the origin of the continuous trajectory, determined by pre-processing. Second, we assigned locations to individual cell data points on the trajectory. We assigned each cell to the closest MST segment (lines between k means) or vertex (k-mean) with a single perpendicular projection. Third, we straightened all the segments (Figure 2. 6A.d) into one horizontal line, and determined the relative order of the assigned locations of single cell data points on the trajectory. Pseudotime values ranged from 0 (at the origin) to 1 (at the end) (Figure 2. 6.d).

Waterfall 3 - Gene expression analysis by Hidden Markov model: We used custom R codes to apply a Hidden Markov model (HMM) to predict gene expression states throughout pseudotime. Briefly, we divided pseudotime into 40 bins, each of which contained an average of 2.5 single cells. We averaged the expression level within each bin and assigned the expression values to observed variables for HMM. We used Baum-Welch algorithm to extract the most probable emission probabilities and transition probabilities. Using the output from Baum-Welch algorithm along with observed variables, we applied the Viterbi algorithm to predict binary gene expression states (Figure 2. 6B).

Functional gene expression analysis: We calculated the Spearman correlation coefficient between pseudotime points and each gene's expression TPM values. Genes with relatively high Spearman correlations were defined as UP genes and genes with relatively low correlations were DOWN genes and the highest and lowest 1,000 genes defined as UP^{1000} and $DOWN^{1000}$, respectively (Table 3. 1). A small subset of the UP^{1000} and $DOWN^{1000}$ genes with low average expression values (< 50 TPM) and low coefficient of variation (< 1.95) were from repeat elements within their exons and excluded from downstream analyses. Raw mapping profiles of all genes shown in pseudotime figures were closely inspected to rule out false positives. We identified transcription factors using public databases (Zhang et al., 2012). We used GO (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) and

wikipathway (Pico et al., 2008) for functional gene ontology analyses, along with the R bioconductor package (Gentleman et al., 2004) or Cytoscape software (Shannon et al., 2003).

For alternative functional gene expression analysis (Figure 3. 8), we first divided the entire transcriptome dataset into three equal groups based on their Spearman correlation to pseudotime: positively correlated, uncorrelated, and negatively correlated. We then evaluated the proportion of positively correlated genes versus negatively correlated genes within each functional entity from an independent functional annotation database wikipathway rather than the KEGG pathway database. If a functional entity contained a disproportionally larger number of up-regulated genes than down-regulated genes, we considered the functional entity to be generally activated, and, conversely, a disproportionally larger number of down-regulated genes indicated that the pathway was generally inactivated over time.

Validation with in situ Database Comparison, Immunohistology, Genetic Labeling and Electrophysiology

We validated our NPC-enriched genes by the Allen mouse brain atlas in situ hybridization dataset (Lein et al., 2007) (Table 2. 3). We inspected the gene expression patterns within the adult dentate gyrus at sagittal views. Genes with clear and relatively even distribution within the SGZ were determined to be “SGZ enriched”, whereas genes with subtle or scattered enrichment within the SGZ were determined to be “ambiguous”. Genes without any enrichment at the SGZ were determined to be “not SGZ enriched”.

Adult *Nestin-GFP*^{cyto} animals (Encinas et al., 2011) were used for immunohistochemical validation. *Hopx-CreER*$^{T2\ f/+}$*;;mT/mG*$^{f/+}$ mice were generated by crossing *Hopx-CreER*$^{T2\ f/+}$ (Takeda et al., 2011) (Strain: Hopx$^{tm2.1(cre/ERT2)Joe}$/J, Jackson Labs Stock: 017606) with the *mT/mG*$^{f/f}$ reporter line (Strain: B6.129(Cg)-Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J; Jackson Labs Stock: 007676). Tamoxifen (62 mg/ml; Sigma; T5648) was prepared in a 5:1 ratio of corn oil/ethanol and heated to 37°C and mixed. Eight week-old *HopX-CreER*$^{T2\ f/+}$*;;mT/mG*$^{f/+}$ animals were injected intraperitoneally with 124 mg/kg tamoxifen. Animals were perfused and extracted brains were placed in 4% paraformaldehyde overnight at 4°C, and then 30% sucrose for 48 hours at 4°C before coronal sections (45 μm) were cut. Immunohistology was performed using antibodies as previously described (Bonaguidi et al., 2011). The following antibodies were used: Aldoc (1:200, goat;

Cat#SC12065; Santa Cruz), GFAP (1:2000, rabbit; Cat#Z0334; DAKO), GFP (1:1000, chicken; Cat#GFP-1020; Aves), GFP (1:1000, goat; Cat#600-101-215; Rockland), Nestin (1:500, chicken; Cat#NES; Aves), PCNA (1:2000, rabbit; Cat#ab18197; Abcam), PCNA (1:500, goat; Cat#SC9857; Santa Cruz), Stmn1 (1:200, rabbit; Cat#ab24445; Abcam), Tbr2 (1:1000, rabbit Cat#Ab23345; Abcam). GFP cells were identified with an Axiovert 200M microscope (Zeiss) and then acquired as z-stacks on Zeiss 710 single-photon confocal microscope using 40X or 63X objectives. For quantification of Stmn1, Aldoc and PCNA expression in *Nestin-GFP*^{cyto} mice, Z-stacks were acquired from 3 animals. Images were analyzed using Imaris 7.1.1 (Bitplane). RGLs were identified by their radial process and soma situated in the SGZ and IPCs were identified by their small soma and tangential process as previously described (Bonaguidi et al., 2011).

Adult *nestin-GFP*^{cyto} transgenic mice were used to validate expression of functional glutamate receptors on NSCs. GFP^{+} radial glia like NSCs in slices prepared acutely from adult animals were recorded by whole-cell patch-clamp upon puffing of AMPA or NMDA in the presence or absence of antagonists as previously described (Song et al., 2012).

References

Arnoldini, M., et al. (2014). "Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation." PLoS Biol **12**(8): e1001928.

Bendall, S. C., et al. (2014). "Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development." Cell **157**(3): 714-725.

Bendall, S. C., et al. (2011). "Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum." Science **332**(6030): 687-696.

Berg, D. A., et al. (2015). "Tbr2-expressing intermediate progenitor cells in the adult mouse hippocampus are unipotent neuronal precursors with limited amplification capacity under homeostasis." Frontiers in Biology **in press**.

Bodenmiller, B., et al. (2012). "Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators." Nat Biotechnol **30**(9): 858-867.

Bonaguidi, M. A., et al. (2011). "In vivo clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics." Cell **145**(7): 1142-1155.

Bracko, O., et al. (2012). "Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis." J Neurosci **32**(10): 3376-3387.

Buettner, F., et al. (2015). "Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells." Nat Biotechnol.

Cahan, P., et al. (2014). "CellNet: network biology applied to stem cell engineering." Cell **158**(4): 903-915.

Codega, P., et al. (2014). "Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche." Neuron **82**(3): 545-559.

Encinas, J. M., et al. (2011). "Division-coupled astrocytic differentiation and age-related depletion of neural stem cells in the adult hippocampus." Cell Stem Cell **8**(5): 566-579.

Encinas, J. M., et al. (2006). "Fluoxetine targets early progenitor cells in the adult brain." Proc Natl Acad Sci U S A **103**(21): 8233-8238.

Guo, G., et al. (2013). "Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire." Cell Stem Cell **13**(4): 492-505.

Hoppe, P. S., et al. (2014). "Single-cell technologies sharpen up mammalian stem cell research." Nat Cell Biol **16**(10): 919-927.

Knobloch, M., et al. (2013). "Metabolic control of adult neural stem cell activity by Fasn-dependent lipogenesis." Nature **493**(7431): 226-230.

Kriegstein, A. and A. Alvarez-Buylla (2009). "The glial nature of embryonic and adult neural stem cells." Annu Rev Neurosci **32**: 149-184.

Levsky, J. M., et al. (2002). "Single-cell gene expression profiling." Science **297**(5582): 836-840.

Li, B. and C. N. Dewey (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." BMC Bioinformatics **12**: 323.

Li, L. and H. Clevers (2010). "Coexistence of quiescent and active adult stem cells in mammals." Science **327**(5965): 542-545.

Lim, H. N. and A. van Oudenaarden (2007). "A multistep epigenetic switch enables the stable inheritance of DNA methylation states." Nat Genet **39**(2): 269-275.

Lu, R., et al. (2011). "Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding." Nat Biotechnol **29**(10): 928-933.

Ming, G. L. and H. Song (2011). "Adult neurogenesis in the mammalian brain: significant answers and significant questions." Neuron **70**(4): 687-702.

Muramoto, T., et al. (2012). "Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation." Proc Natl Acad Sci U S A **109**(19): 7350-7355.

Novick, A. and M. Weiner (1957). "Enzyme Induction as an All-or-None Phenomenon." Proc Natl Acad Sci U S A **43**(7): 553-566.

Ozbudak, E. M., et al. (2004). "Multistability in the lactose utilization network of Escherichia coli." Nature **427**(6976): 737-740.

Pico, A. R., et al. (2008). "WikiPathways: pathway editing for the people." PLoS Biol **6**(7): e184.

Qiu, P., et al. (2011). "Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE." Nat Biotechnol **29**(10): 886-891.

Raj, A., et al. (2006). "Stochastic mRNA synthesis in mammalian cells." PLoS Biol **4**(10): e309.

Ramskold, D., et al. (2012). "Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells." Nat Biotechnol **30**(8): 777-782.

Saadatpour, A., et al. (2014). "Characterizing heterogeneity in leukemic cells using single-cell gene expression analysis." Genome Biol **15**(12): 525.

Schofield, R. (1978). "The relationship between the spleen colony-forming cell and the haemopoietic stem cell." Blood Cells **4**(1-2): 7-25.

Schwanhausser, B., et al. (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.

Shalek, A. K., et al. (2013). "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." Nature **498**(7453): 236-240.

Signer, R. A., et al. (2014). "Haematopoietic stem cells require a highly regulated protein synthesis rate." Nature **509**(7498): 49-54.

Simons, B. D. and H. Clevers (2011). "Strategies for homeostatic stem cell self-renewal in adult tissues." Cell **145**(6): 851-862.

Stegle, O., et al. (2015). "Computational and analytical challenges in single-cell transcriptomics." Nat Rev Genet **16**(3): 133-145.

Takeda, N., et al. (2011). "Interconversion between intestinal stem cell populations in distinct niches." Science **334**(6061): 1420-1424.

Thorndike, R. L. (1953). "Who belongs in the family?" Psychometrika **18**(4): 267-276.

Trapnell, C., et al. (2014). "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells." Nat Biotechnol **32**(4): 381-386.

Trapnell, C., et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Treutlein, B., et al. (2014). "Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq." Nature **509**(7500): 371-375.

Usoskin, D., et al. (2015). "Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing." Nat Neurosci **18**(1): 145-153.

Wagner, G. P., et al. (2012). "Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples." Theory Biosci **131**(4): 281-285.

Curriculum Vitae

JAEHOON SHIN

Graduate student in the laboratory of Hongjun Song

Institute for Cell Engineering

Johns Hopkins Medical Institutions

Phone: (443) 676-5582 · E-mail: shin@jhmi.edu

Education

BS in Chemical Engineering (*summa cum laude*), Seoul National University, **2000 - 04**

MD, Seoul National University, **2004 - 08**

PhD, Johns Hopkins Medical Institutions, **2010 - current**

Scholarship, Awards, Honors

1. Honored Student, Department of Chemical Engineering, Seoul National University, Korea, **2001-04**

2. General Electric - Fulbright Scholarship, Korea, **2002-04**

3. Alumni prize, Department of Chemical Engineering, Seoul National University, Korea, **2004**

4. Doosan Scholarship, Yeonkang Foundation, Korea, **2006-08**

5. Samsung Scholarship, Samsung Foundation of Culture, Korea, **2010-2015**

6. Finalist, Innocentive challenge for single cell analysis, National Institute of Health (NIH), USA, **2015**
   - *NIH sponsored national competition (http://commonfund.nih.gov/singlecell/challenge)*

Publications

1. Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G.E., Nauen, D.W., Christian, K.M., Ming, G.-l., Song, H. (2015). Single-cell RNA-seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell 17(3), 360-72. *(featured as a cover)*
2. Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J., Huang, L., Wan J., Wu H., Qian J., Cheng X., Zhu H., Ming, G.-l., Jin P., Song H. (2015). Lin28A Binds Active Promoters and Recruits Tet1 to Regulate Gene Expression. Molecular cell 61, 1 153-1160.
3. Shin, J., Ming, G.-l., Song, H. (2015). Molecular Toggle Switch of Histone Demethylase LSD1. Molecular Cell 57, 949-950.
4. Shin, J., Ming, G.-l., Song, H. (2015). Seeking a Roadmap toward Neuroepigenetics. Neuron 86, 12-15.
5. Yu, H., Su, Y., Shin, J., Zhong, C., Guo, J.U., Weng, Y.-L., Gao, F., Geschwind, D.H., Coppola, G., Ming, G.-l. (2015). Tet3 regulates synaptic transmission and homeostatic plasticity via DNA oxidation and repair. Nature Neuroscience.
6. Shin, J., Ming, G.-l., Song, H. (2014). Decoding neural transcriptomes and epigenomes via high-throughput sequencing. Nature Neuroscience, 17(11), 1463-1475.
7. Wen, Z., Nguyen, H.N., Guo, Z., Lalli, M. A., Wang, X., Su, Y., Kim, N.-S., Yoon, K.-J., Shin, J., Zhang, C., Makri, G., Nauen, D.W., Yu, H., Guzman, E., Chiang, C.H., Yoritomo, N., Kaibuchi, K., Zou, J., Christian, K.M., Cheng, L., Ross, C.A., Margolis R.L., Chen, G., Kosik, K.S., Song, H., Ming, G.-l. (2014). Synaptic dysregulation in a human iPS cell model of mental disorders. Nature, 515(7527), 414-418.
8. Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G.-l., Song, H. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nature Neuroscience, 17(2), 215-222.
9. Shin, J., Ming, G.-l., Song, H. (2014). DNA modifications in the mammalian brain. Philosophical

Transactions of the Royal Society B: Biological Sciences, 369(1652), 20130512.
10. Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N., Shin, J., Cox, E., Rho, H.S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G.-l., Wade, H., Song, H., Qian, J., Zhu, H. (2013). DNA Methylation Presents Distinct Binding Sites for Human Transcription Factors. eLife, 2, e00726. doi:10.7554/eLife.00726.
11. Weng, Y.L., An, R., Shin, J., Song, H., Ming, G.-l. (2013). DNA modifications and neurological disorders. Neurotherapeutics, 10(4), 556-567.
12. Shin, J., Ming, G.-l., Song, H. (2013). By Hook or by Crook: Multifaceted DNA-Binding Properties of MeCP2. Cell, 152(5), 940-942.
13. Bhang, S.H., Park, J., Yang, H.S., Shin, J., Kim, B.-S. (2013). Platelet-rich plasma enhances the dermal regeneration efficacy of human adipose-derived stromal cells administered to skin wounds. Cell Transplantation, 22(3), 437-445.
14. Yoon, H.H., Bhang, S.H., Shin, J.Y., Shin, J., Kim, B.-S. (2012). Enhanced cartilage formation via three-dimensional cell engineering of human adipose-derived stem cells. Tissue Engineering Part A, 18(19-20), 1949-1956.
15. *Yang, H.S., *Shin, J., Bhang, S.H., Shin, J.Y., Park, J., Im, G.I., Kim, C.S., Kim, B.-S. (2011). Enhanced skin wound healing by a sustained release of growth factors contained in platelet-rich plasma. Experimental and Molecular Medicine, 43(11), 622-629. (* equal contributions)

Presentations

Oral presentation

1. Shin, J. (2014). Single Cell RNA-sequencing Reveals Molecular Dynamics of Adult Hippocampal Neural Stem Cells, Oral Presentation presented at: Gordon Research Conference; Boston, MA

Poster presentation

1. Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Nauen, D.W., Song, J., Bonaguidi, M.A., Ming, G.-l., Song, H. (2014). Single Cell RNA-seq Reveals Molecular Dynamics of Adult Hippocampal Neural Stem Cells, Poster presented at: Howard Hughes Medical Institute, Janelia Farm Conference; Ashburn, MD